PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
A23J 1/00, C07K 3/00
C12Q 1/68, G01N 1/00

A1

(11) International Publication Number: WO 93/08701

(43) International Publication Date: 13 May 1993 (13.05.93)

(21) International Application Number: PCT/US92/08603

(22) International Filing Date: 9 October 1992 (09.10.92)

(30) Priority data:
07/785,567 30 October 1991 (30.10.91) US

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; Fruit Street, Boston, MA 02114 (US).

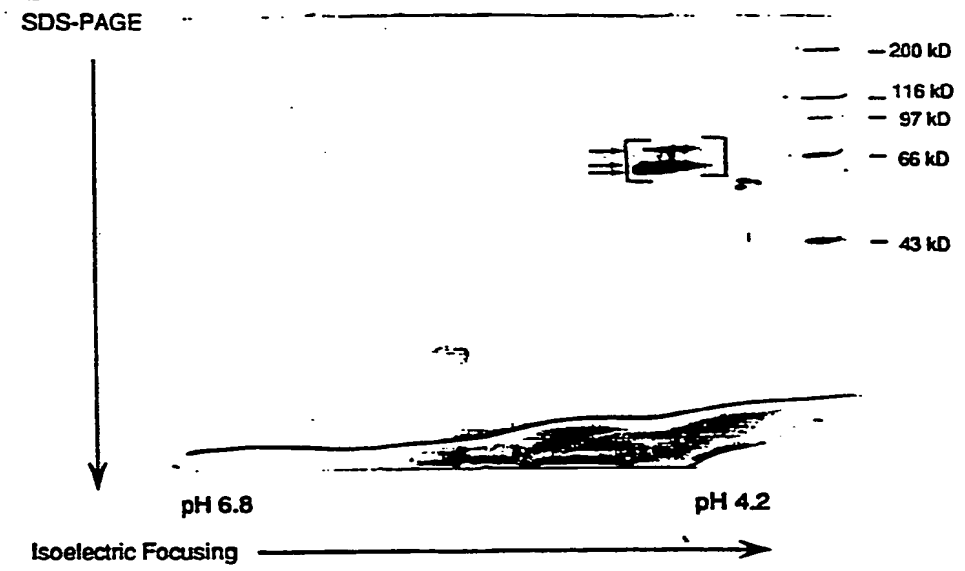
(72) Inventors: KINGSTON, Robert, E. ; 32 Draper Avenue, Arlington, MA 02174 (US). PAPOULAS, Ophelia ; 10 Emerson Place (15-B), Boston, MA 02114 (US).

(74) Agents: GOLDSTEIN, Jorge, A. et al.; Sterne, Kessler, Goldstein & Fox, 1225 Conneticut Avenue, N.W., Suite 300, Washington, DC 20036 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: C-MYC DNA BINDING PARTNERS, MOTIFS, SCREENING ASSAYS, AND USES THEREOF



(57) Abstract

The development of a mammalian cell line that overexpresses Myc and the purification of significant quantities of c-Myc from these cells is described. Three types of c-Myc-driven protein oligomerization (or complex) formations are described: (1) homo-oligomer complexes (herein termed C1 complexes) formed by association of at least two peptides of c-Myc, (2) hetero-oligomer complexes (herein termed C2 complexes) formed by heterodimerization of at least two peptides, at least one of which is not the c-Myc peptide, and specifically hetero-oligomerization between c-Myc and a 26-29 kd factor, and (3) c-Myc-dependent hetero-oligomeric complexes (herein termed the C2' complex) formed in the presence of c-Myc, however such hetero-oligomeric proteins not containing any peptides which are c-Myc. The invention is further directed to a reliable and accurate method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SK	Slovak Republic
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
UA	Ukraine
US	United States of America
VN	Viet Nam

TITLE OF THE INVENTION

C-MYC DNA BINDING PARTNERS, MOTIFS, SCREENING ASSAYS, AND USES THEREOF

Cross-Reference to Related Applications

This application is a continuation-in-part of U.S. Application No. 07/510,253, filed April 19, 1990.

Field of the Invention

This invention is directed to methods for the purification of mammalian Myc protein, and methods for the identification of compounds that inhibit c-Myc transcriptional activity.

BACKGROUND OF THE INVENTION

Myc is a nuclear oncogene whose aberrant expression is associated with many different types of human cancers in many different tissues (Cole, M.D., *Ann. Rev. Genet. 20*:361-384 (1986)). While the mechanism of c-Myc oncoprotein action remains unknown, it clearly plays a role in the control of cell growth and differentiation (Lüscher and Eisenman, *Genes & Dev. 4*:2025-2035 (1990); Penn *et al.*, *Sem. Cancer Biol. 1*:69 (1990)). One plausible mechanism of Myc action is as a regulator of transcription in a pathway directly controlling proliferation and differentiation. This model is consistent with several observations. First, Myc has long been known as a nuclear protein with a general affinity for DNA (Abrams *et al.*, *Cell 29*:427-439 (1982); Alitalo *et al.*, *Nature 306*:274-277 (1983); Donner *et al.*, *Nature 296*:262-265 (1982); Persson and Leder, *Science 225*:718-721 (1984)), and recently a site has been identified which is specifically bound by bacterially expressed variants of c-Myc (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Prendergast and Ziff, *Science 251*:186-189 (1991)). Second, full length c-Myc has been shown to both activate and repress genes in transient transfection assays (Kaddurah-Daouk *et al.*, *Genes & Dev. 1*:347-357 (1987); Yang *et al.*, *Mol. Cell. Biol. 11*:2291-2295

(1991)), and will weakly stimulate transcription when fused to a heterologous DNA-binding domain (Lech *et al.*, *Cell 52*:179-184 (1988); Kato *et al.*, *Mol. Cell. Biol. 10*:5914-5920 (1990)). And finally, sequence similarities described below place Myc in the company of known transcription factors.

Myc contains two domains that suggest it oligomerizes, perhaps as a dimer, and binds specifically to DNA: a leucine zipper domain and a basic-helix-loop-helix (B-HLH) domain. The leucine zipper is an α-helical structure found in sequence specific DNA-binding proteins such as Fos and Jun where it mediates homo- or heterodimerization via a coiled-coiled interaction (Landschulz *et al.*, *Science 240*:1759-1764 (1988); O'Shea *et al.*, *Science 243*:538-542 (1989); and reviewed in Busch and Sassone-Corsi, *TIG 6*:36-40 (1990)). This dimerization is necessary for DNA binding (Gentz *et al.*, *Science 243*:1695-1699 (1989); Halazonetis *et al.*, *Cell 55*:917-924 (1988); Kouzarides and Ziff, *Nature 336*:646-651 (1988); Turner and Tjian, *Science 243*:1689-1694 (1989)). The HLH region also appears to mediate oligomerization necessary for DNA binding in several developmentally important proteins (Murre *et al.*, *Cell 58*:537-544 (1989); Murre *et al.*, *Cell 56*:777-783 (1989)). HLH proteins form a large and growing family and include the products of the *achaete-scute* and *daughterless* genes responsible for neural development in *Drosophila*, the R gene family which regulates pigment pattern in corn, MyoD and several other proteins involved in muscle specific differentiation in vertebrates, and a centromere binding protein, CBF1, from yeast (Braun *et al.*, *EMBO J. 8*:701-709 (1989); Cai and Davis, *Cell 61*:437-446 (1990); Caudy *et al.*, *Cell 55*:1061-1067 (1988); Cronmiller *et al.*, *Genes & Dev. 2*:1666-1676 (1988); Davis *et al.*, *Cell 51*:987-1000 (1987); Edmondson and Olson, *Genes & Dev. 3*:628-640 (1989); Ludwig and Wessler, *Cell 62*:849-851 (1990); Pinney *et al.*, *Cell 53*:781-793 (1988); Rhodes and Konieczny, *Genes & Dev. 3*:2050-2061 (1989); Villares and Cabrera, *Cell 50*:415-424 (1987); Wright *et al.*, *Cell 56*:607-617 (1989)). While many proteins

contain either an HLH or leucine zipper motif, Myc is one of a smaller number of proteins which contain both an HLH and a leucine zipper. Both the leucine zipper containing proteins and the HLH proteins require a stretch of basic amino acids adjacent to the dimerization motif to contact DNA (reviewed in Busch and Sassone-Corsi, *TIG 6*:36-40 (1990); Jones, N., *Cell 61*:9-11 (1990)). Interestingly, all B-HLH proteins appear to bind to closely related DNA sequences known as E-Boxes. These are sequence motifs found in the immunoglobulin and other tissue specific enhancers having a core of NNCANNTGNN [SEQ ID No. 16] where different central bases are preferred by different B-HLH proteins and the flanking bases can affect binding affinity (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Blackwell and Weintraub, *Science 250*:1104-1110 (1990)). The core of the reported binding site for c-Myc, CACGTG, fits this pattern and has the same core sequence as the upstream sequence element (USE) of the Adenovirus major late promoter (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Prendergast and Ziff, *Science 251*:186-189 (1991)). A cellular transcription factor (USF or MLTF) which binds to the USE has recently been cloned and also contains a B-HLH domain adjacent to a leucine zipper (Gregor *et al.*, *Genes & Dev. 4*:1730-1740 (1990)).

Many of these B-HLH or leucine zipper proteins have been found to form not only homodimers but heterodimers with other proteins having like dimerization motifs (reviewed in Busch and Sassone-Corsi, *TIG 6*:36-40 (1990); Jones, N., *Cell 61*:9-11 (1990)). Heterodimerization between specific groups of B-HLH or leucine zipper proteins can alter their DNA binding properties. While homodimers might bind weakly, heterodimers with the appropriate partner can bind with increased affinity and in some cases with a new specificity (Jones, N., *Cell 61*:9-11 (1990); Blackwell and Weintraub, *Science 250*:1104-1110 (1990); Wright *et al.*, *Mol. Cell. Biol. 11*:4104-4110 (1991)). Myc is capable of forming a homo-oligomer at high concentrations *in vitro* (Dang *et al.*, *Nature 337*:664-666 (1989); Kerkhoff and Bister, *Oncogene 6*:93-102 (1991)), although it is not clear whether

that homo-oligomer actually forms *in vivo* (Dang *et al.*, *Mol. Cell. Biol. 11*:954-962 (1991)). It seems likely that Myc directly interacts with other cellular protein(s) to form hetero-oligomer(s), and indeed one such "partner" protein, designated Max, has recently been identified (Blackwood and Eisenmann, *Science 251*:1211-1217 (1991)). The effect that such partner proteins have on Myc DNA-binding specificity is likely to be central to understanding the function of Myc.

Much of the *in vitro* work done on B-HLH proteins has utilized *in vitro* transcribed and translated protein or has used protein overexpressed in bacteria. Myc expressed by these means has been used to determine binding specificity and to demonstrate that Myc can form heterodimers with Max (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Prendergast and Ziff, *Science 251*:186-189 (1991); Blackwood and Eisenmann, *Science 251*:1211-1217 (1991)). Myc, however, is post-translationally modified by at least phosphorylation in mammalian cells (Hann and Eisenmann, *Mol. Cell. Biol. 4*:2486-2497 (1984); Ramsay *et al.*, *Proc. Natl. Acad. Sci. USA 81*:7742-7746 (1984)), and post-translational modifications are believed to regulate the function of many proteins, including the transcription factors Myb, Fos, HSF, CREB, and SP-1 (Abate *et al.*, *Science 249*:1157-1161 (1990); Jackson *et al.*, *Cell 63*:155-165 (1990); Lüscher *et al.*, *Nature 344*:517-522 (1990); Sorger *et al.*, *Nature 329*:81-84 (1987); Yamamoto *et al.*, *Nature 334*:494-498 (1988)). In addition, Myc produced in avian cells has been reported to bind more tightly to DNA cellulose than bacterially produced Myc (Kerkhoff and Bister, *Oncogene 6*:93-102 (1991)).

Several lines of evidence argue that the biochemical function(s) of Myc will be determined in large part by hetero-oligomerization with Max and perhaps with other, as yet unidentified, factors. A complete understanding of the function of c-Myc will therefore require the identification of all partner proteins and a functional characterization of the complexes that these proteins form in the absence or presence of c-Myc.

To unravel the nature of Myc's function it will be necessary to determine not only the binding properties of all relevant complexes but to ascertain how they differ in action once bound. Post-translational modification might play a role in modulating the formation, binding, or further activities of these complexes and the availability of large quantities of modified c-Myc, such as described here, should facilitate a biochemical approach to this problem. Such studies should lead us to an understanding of the complexes available at different times in different cell types and the consequences for each cell in terms of appropriate growth and differentiation, or oncogenesis.

Further, to date, no inhibitors of c-Myc action have been identified. The identification of such inhibitors has suffered for lack of identification of a specific DNA binding sequence to which c-Myc binds, and for lack of a simple, inexpensive and reliable screening assay which could rapidly identify potential inhibitors and active derivatives thereof. Thus a need also still exists for rapid, economical screening assays which identify specific inhibitors of c-Myc activity.

SUMMARY OF THE INVENTION

Recognizing the potential importance of inhibitors of c-Myc oncoprotein activity in the therapeutic treatment of many forms of cancer, and cognizant of the lack of a simple assay system in which such inhibitors might be identified, the inventors have investigated c-myc DNA binding.

These efforts led to the development of a mammalian cell line that overexpresses Myc and the purification of significant quantities of c-Myc from these cells. These efforts culminated in the discovery of three types of c-Myc-driven protein oligomerization (or complex) formations: (1) homo-oligomer complexes (herein termed C1 complexes) formed by association of at least two peptides of c-Myc, (2) hetero-oligomer complexes (herein termed C2 complexes) formed by heterodimerization of

at least two peptides, at least one of which is not the c-Myc peptide, and specifically hetero-oligomerization between c-Myc and a 26-29 kd factor, and (3) c-Myc-dependent hetero-oligomeric complexes (herein termed the C2´ complex) formed in the presence of c-Myc, however such hetero-oligomeric proteins not containing any peptides which are c-Myc.

Accordingly, the invention is directed to a reliable and accurate method for the purification of Myc from a mammalian source.

The invention is further directed to the use of oligomers containing the DNA motif 5'-CACGTG-3', in its double stranded DNA form, as a reliable and accurate method for the detection of the presence of C1 complexes in a sample.

The invention is further directed to the use of the DNA motif 5'-CACGTG-3', in its double stranded DNA form, as a reliable and accurate method for the detection of C2 complexes in a sample.

The invention is further directed to the use of the DNA motif 5'-CAGCTG-3', in its double stranded DNA form, as a reliable and accurate method for the detection of C2´ complexes in a sample.

The invention is further directed to a 26-29 kD protein fraction purified from Chinese hamster ovary (CHO) cells or baculovirus, such protein fraction containing at least one peptide capable of forming C2 complex oligomers with c-Myc.

The invention is further directed to a 40-50 kD protein fraction purified from CHO cells, such protein fraction containing at least one peptide capable of forming C2´ complex oligomers in the presence of c-Myc.

The invention is further directed to a reliable and accurate method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity, and especially as an inhibitor of C1 complex formation, C2 complex formation or C2´ complex formation.

The invention is further directed to a reliable and accurate method for objectively classifying compounds, including human pharmaceuticals, as

inhibitors of c-Myc activity, and especially as an inhibitor of C1 complex DNA binding, C2 complex DNA binding, or C2′ complex DNA binding.

The invention further provides a method for identifying and classifying the mechanism of action of a bioactive c-Myc-inhibiting compound.

The invention further provides an assay for the monitoring of the isolation and/or purification of a peptide capable of forming a C2 or C2′ complex, or a mixture of such peptides from a crude preparation.

The invention further provides an assay for the monitoring of the isolation and/or purification of an c-Myc-inhibiting compound or mixture of such compounds from a crude preparation of such compounds.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. Purified c-Myc Protein. A) 1 μg of c-Myc protein purified from the 5A overexpressing CHO cell line was subjected to 2-dimensional gel electrophoresis. An isoelectric focusing tube gel was run with pH 5-7 ampholytes followed by SDS-PAGE and silver staining. The Myc proteins are bracketed and arrows distinguish the 60, 62, and 72 kD species. The gel was trimmed for this figure; the actual pI range for the Myc proteins was 5.0-5.6. B) 0.5 μg of purified c-Myc protein from the indicated cell lines was electrophoresed on an SDS gel and either visualized by silver staining (left lane) or electroblotted to nitrocellulose and subjected to immunoblotting using the ST-2 polyclonal antibody (right 2 lanes).

Fig. 2. DNA Binding of Purified c-Myc Proteins. The EMSA was carried out as described in materials and methods using equal amounts (approximately 2 ng) of the following probes and 0.5 μg of either purified CHO produced c-Myc or baculovirus produced c-Myc: $(\mu E2)_3$ lanes 1 and 7, $(\mu E3)_3$ lanes 2 and 8, MLC-A lanes 3 and 9, MLC-B lanes 4 and 10, $(USE)_3$ lanes 5, 11, and 12, and HSE lanes 6 and 13. Full probe

sequences are given in materials and methods. Lanes 1-6 and lanes 7-13 are different exposures of lanes from the same gel.

Fig. 3. C1 Binding Activity is Present in Myc containing Slices of SDS Gels. 400 μg of CHO produced c-Myc or 163 μg of baculovirus produced c-Myc was separated on an SDS-PAGE gel. Proteins from 0.5 cm slices were recovered, renatured as described in materials and methods, and analyzed by EMSA using the (USE)3 probe. 0.4 μg of the CHO Myc load and 5 μl of the protein from the CHO Myc-containing slice were analyzed on the same gel (left panel). 0.37 μg of the baculovirus Myc load and 5 μl of the protein from the baculovirus Myc slice were analyzed on the same gel (right panel). Slices from other molecular weight ranges of the same gel showed no binding (data not shown).

Fig. 4. Activity is Formed by c-Myc and a 26-29 kD Factor. Proteins from gel slices were recovered and analyzed by EMSA as described in materials and methods using the $(USE)_3$ probe. Lanes 1-4 represent proteins from the same gel loaded with baculovirus produced Myc described for Fig. 5. These lanes contain 0.37 μg of the loaded material (lane 1), 0.75 μg BSA with 7.5 μl of proteins from either a Myc slice (lane 2) or a 26-29 kD slice (lane 3), or 7.5 of each slice used for lanes 1 and 2 plus 0.2 μg of BSA (lane 4). Lanes 5-8 and 10 contain proteins from gels loaded with Myc purified from CHO cells. These lanes contain 0.47 of the gel load (lane 5), 4 μl of material from a Myc slice of a gel loaded with 400 μg of Myc (lane 6), 7 μl of material from a 26-29 kD slice of a similar gel plus 0.8 μg Protein A (lane 7), and both 4 μl of the Myc slice and 7 μl of the 26-29 kD slice (lane 8). Lanes 9-12 utilize the bacterially expressed Protein A-Myc fusion proteins containing either the Myc B-HLH and leucine zipper domains (amino acids 353-439) or lacking the basic region and containing Myc amino acids 372-439. These were expressed and purified as described in materials and methods. Lane 9 contains 0.5 μg of Protein A-Myc(353-439) and lane 10 contains the same plus 7 μl of the 26-29 kD slice. Lane 11 contains 1 μg of Protein A-

Myc(372-439) and lane 12 contains 0.5 μg of Protein A-Myc(372-439) plus 7 μl of the 26-29 kD slice.

Fig. 5. C2' Binding Activity Requires a 40-50 Kd Factor. A) 101 μg of CHO produced c-Myc was separated on an SDS gel. Proteins were recovered, resuspended in 100 μl, and renatured and analyzed by EMSA using the ERP3/4 probe. This probe contains the portion of the MLC enhancer that encompasses the μE2 site. EMSA samples contained 0.3 μg of the SDS gel load (lane 1), 7.5 μl of the proteins from the Myc slice (lane 2), or the 40-50 kD slice (lane 3), or 7.5 μl of both slices renatured together (lane 4). B) EMSA samples contained 0.9 μg purified baculovirus produced c-Myc (lane 5), 3 μl of protein from the 40-50 kD slice of a gel loaded with 400 μg CHO produced c-Myc (lane 6), or both renatured together (lane 7). The probe was ERP1/2. C) EMSA samples contained 10 μl (0.9 μg) of bacterially produced c-Myc fusion protein containing Myc amino acids 353-439 (lane 8), 0.47 μg of CHO produced c-Myc (lane 9), 5 μl of protein from the 40-50 kD slice of a gel loaded with 400 μg of the CHO Myc shown in lane 9 (lane 10), or 5 μl of the same 40-50 kD material renatured in the presence of either 0.9 μg of the baculovirus produced Myc shown in lane 5 (lane 11), 2 μl (0.18 μg) of the bacterially produced Myc fusion protein containing Myc amino acids 353-439 (lane 12), or 4 μl (0.36 μg) of the same bacterially produced Myc fusion protein (lane 13). The probe was ERP1/2.

Fig. 6. Antibodies to c-Myc Interact with the C1 and C2 Complexes. EMSA reactions were set up with the indicated Myc protein preparations (0.37 μg baculovirus produced c-Myc or 0.47 μg of CHO produced c-Myc). These reactions were preincubated 30 min on ice in the presence of the indicated antibody (α-Myc monoclonal 1F7 or a monoclonal directed against the lambda repressor, cI). 1 ng of SMS probe or μE2-containing probe number 7 (see Fig. 7) was added subsequently and binding and electrophoresis were as usual.

Fig. 7. Oligonucleotides Selected from Random Sequence after 8 Rounds of EMSA. Sequences were selected from oligonucleotides containing 20 base pairs of random sequence using a reiterative EMSA procedure described in materials and methods. Underlined nucleotides are from the PCR primer sites. Tables below the aligned sequences tabulate the frequency of each base in the 6 flanking positions surrounding the CACGTG motifs. Only bases next to a perfect fit of the CACGTG core were tabulated since sequences without this core were found not to function as high affinity binding sites (Fig. 8, and data not shown). Bold numbers adjacent to individual sequences indicate those oligonucleotides which were tested individually by EMSA in Fig. 8. Asterisks indicate additional sequences which were tested individually (data not shown).

Fig. 8. Selected Sites form Predicted Complexes. EMSA was carried out using either 2.8 ng of the SMS probe or equal amounts (1 ng) of probes 1-11 indicated in Fig. 7. Probes 1-11 were labeled and gel isolated in parallel and had approximately equal specific activities. Binding reactions contained either no additional protein (-), 0.37 μg of baculovirus produced c-Myc (B) or 0.47 μg of CHO produced c-Myc (C). Free probe is visible at the bottom of the gel.

Fig. 9. Off-Rate of the C1 and C2 Complexes. The standard EMSA reaction was scaled up for 11 samples containing 0.4 μg of purified baculovirus produced c-Myc per sample. Probe and competitor were $(USE)_3$. After allowing 20 min for binding 20 μ1 was loaded on a prerun EMSA gel as a measure of the starting amount of complex (ST) and enough cold competitor was added to the remaining sample to achieve a 250 fold molar excess over probe. Immediately upon addition of competitor the sample was gently mixed and 20 μ1 aliquots were loaded at the indicated times (0, 30 s, 1 min, 4 min, etc.). A control sample (C) was made up individually in which competitor was added prior to the start of binding to demonstrate complete competition. This sample was loaded at the same time as the ST sample. All samples were loaded on a

continuously running gel so that the complex in the starting lane runs ahead of the equivalent complex in lanes loaded later.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the description that follows, a number of terms used in recombinant DNA technology are extensively utilized. In order to provide a clearer and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

Oligomer of Interest. As used herein, an "oligomer of interest" refers to any of the following types of oligomeric proteins: first, Myc-containing oligomers including homo-oligomers of Myc peptides (a C1 complex), and hetero-oligomers containing at least one peptide of Myc and one peptide of a Myc "partner" (a C2 complex); second, oligomers that form in the presence of Myc-containing homo-oligomers or Myc-containing hetero-oligomers but which themselves do not contain the Myc peptide, such oligomers including non-Myc-containing homo-oligomers that associate in the presence of Myc and non-Myc-containing hetero-oligomers that associate in the presence of Myc (a C2′ complex).

Oligomer. An "oligomer" as it refers to proteins, means a protein composed of more than one peptide subunit, such as dimers, trimers, tetramers, etc. Such oligomeric protein may be a homo-oligomer, that is, composed entirely of two or more identical subunits; alternatively, such oligomeric protein may be a hetero-oligomer, that is, composed of at least two different peptides. Oligomers containing three or more peptides may contain more than one copy of a peptide.

C2′ Protein(s). As used herein, for convenience, a "C2′ protein" is a protein or peptide that is a member of the second class of the "oligomers-of-interest," that is, a protein that forms oligomers in the presence of Myc, c-Myc homo-oligomers or Myc-containing hetero-oligomers so as to bind to

a specific DNA sequence, but which does not contain a Myc peptide, such oligomers including non-Myc-containing homo-oligomers that associate in the presence of Myc and non-Myc-containing hetero-oligomers that associate in the presence of Myc.

Operably-linked. As used herein, two macromolecular elements are operably-linked when the two macromolecular elements are physically arranged such that factors which influence the activity of the first element cause the first element to induce an effect on the second element. For example, the transcription of a coding sequence which is operably-linked to a promoter element is induced by factors which "activate" the promoter's activity; transcription of a coding sequence which is operably-linked to a promoter element is inhibited by factors which "repress" the promoter's activity. Thus, a promoter region would be operably-linked to the coding sequence of a protein if transcription of the coding sequence activity was influenced by the activity of the promoter.

Response. As used herein, the term "response" is intended to refer to a change in any parameter which can be used to measure, indicate or otherwise describe c-Myc action or oligomer (homo-oligomer (C1 complex) or hetero-oligomer (C2 complex)) formation, including c-Myc dependent hetero-oligomerization (C2´ complex) formation. The response may be revealed as a physical change (such as a change in phenotype) or, it may be revealed as a molecular change (such as a change in a reaction rate or affinity constant). Detection of the response may be performed by any means appropriate. "Detecting" refers to any method by which such response may be evaluated so as to provide a meaningful indicia of whether the event has occurred.

Compound. The term "compound" is intended to refer to a chemical entity, whether in the solid, liquid, or gaseous phase. The term should be read to include synthetic compounds, natural products and macromolecular entities such as polypeptides, polynucleotides, or lipids,

and also small entities such as neurotransmitters, ligands, hormones or elemental compounds.

Bioactive Compound. The term "bioactive compound" is intended to refer to any compound which induces a detectable or measurable response in the methods of the invention.

Promoter. A "promoter" is a DNA sequence located proximal to the start of transcription at the 5' end of the transcribed sequence. The promoter may contain multiple regulatory elements which interact in modulating transcription of the operably-linked gene.

Expression. Expression is the process by which the information encoded within a gene is revealed. If the gene encodes a protein, expression involves transcription of the DNA into mRNA, the processing of mRNA (if necessary) into a mature mRNA product, and translation of the mature mRNA into protein.

A nucleic acid molecule, such as a DNA or gene is said to be "capable of expressing" a polypeptide if the DNA contains the coding sequences for the polypeptide and expression control sequences which, in the appropriate host environment, provide the ability to transcribe, process and translate the genetic information contained in the DNA into a protein product, and if such expression control sequences are operably-linked to the nucleotide sequence which encodes the polypeptide.

Cloning vehicle. A "cloning vehicle" is any molecular entity that is capable of delivering a nucleic acid sequence into a host cell for cloning purposes. Examples of cloning vehicles include plasmids or phage genomes. A plasmid that can replicate autonomously in the host cell is especially desired. Alternatively, a nucleic acid molecule that can insert into the host cell's chromosomal DNA is especially useful.

Cloning vehicles are often characterized by one or a small number of endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of

the vehicle, and into which DNA may be spliced in order to bring about its replication and cloning.

The cloning vehicle may further contain a marker suitable for use in the identification of cells transformed with the cloning vehicle. Markers, for example, are tetracycline resistance or ampicillin resistance. The word "vector" is sometimes used for "cloning vehicle."

Expression vehicle. An "expression vehicle" is a vehicle or vector similar to a cloning vehicle but is especially designed to provide sequences capable of expressing the cloned gene after transformation into a host.

In an expression vehicle, the gene to be cloned is usually operably-linked to certain control sequences such as promoter sequences. Expression control sequences will vary depending on whether the vector is designed to express the operably-linked gene in a prokaryotic or eukaryotic host and may additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

Host. By "host" is meant any organism that is the recipient of a cloning or expression vehicle.

a. Isolation of c-Myc Protein From Mammalian Cells and Preparation of Fractions Containing C2 and C2´ Complex Binding Activity

Although there have been previous reports of purified Myc protein, the present inventors found that the Mvc protein preparations described therein, and the methods used to isolate that protein, failed to achieve the requisite amount of yield needed to sequence characterize Myc action in mammalian sources. The inventors have overcome this problem and describe, for the first time, a unique and useful method for the isolation of highly purified mammalian c-Myc protein which provides the requisite high degree of quantity of mammalian c-Myc protein needed for the

characterization of c-Myc directed DNa binding and biological action. The inventors have also been able to purify large quantities of Myc from a recombinant insect cell system. The purified Myc protein of the invention exhibits the only known biochemical activity of c-Myc, an ability to bind the sequence CACGTG. As a direct result of the method of the invention for the isolation of c-Myc protein, the inventors were able to identify peptides that naturally associate with c-Myc in a hetero-oligomers, or peptides that naturally associate with each other as a result of the action of c-Myc, such peptides found to be present in certain column chromatography fractions of the c-Myc purification scheme.

Accordingly to the invention, purification of Myc from a mammalian source is preferably achieved utilizing a mammalian cell line that overexpresses either recombinant or non-recombinant c-Myc and is performed completely on ice or equivalent temperatures of 0-5°C, using reagents and buffers at the same temperature. For example, the overexpressing Chinese hamster ovary (CHO) cell line 5A is useful for such purification. In CHO 5A cells, recombinant mouse c-Myc is under the control of a regulatable promoter, and has been integrated and amplified in the genome of the parent CHO cell line for maximum stability and production. Except where otherwise noted, for the methods and assays of the invention, the native or recombinant Myc should include at least the two coding exons of Myc.

After collecting the cells by centrifugation using techniques known in the art, and prior to lysis of the outer cell membrane, the cells should be washed at least once in a low salt neutral buffer such as 0.9% NaCl in 10-50 mM phosphate, pH 7.0-7.5 (phosphate buffered saline, PBS) to remove remaining growth medium.

Lysis of the washed cells is also achieved in a low salt, neutral to mildly acidic lysis buffer, preferably about pH 6.8, containing at least one protease inhibitor, such as aprotinin or phenylmethylsulfonyl fluoride (PMSF), preferably containing a combination of such inhibitors. Salts such

as potassium (in the KCl form) and magnesium (in the $MgCl_2$ form) are also preferably added. In addition, nonionic detergents such as NP40 (0.5% v/v) and Na-deoxycholate (0.1 %) should be added.

Cell outer membrane lysis should be performed under conditions that lyse the host cell without lysing the nucleus, or induce significant leakage from the nuclear membrane. The cells may be allowed to sit for a short period of time, for example, 10 minutes, in the detergent-containing lysis buffer before mechanical intervention is utilized in the lysis step. Mechanical intervention is best performed with a gentle disruption of the detergent treated cells, for example, utilizing 40 strokes in a Dounce homogenizer with a type A pestle, or the equivalent of such treatment.

Nuclei may be collected from the lysed cell preparation using techniques known in the art, such as, for example, centrifugation at 1000xg for 5 min at 4°C and washed at least once in the same low salt lysis buffer used to lyse the outer cell membrane.

Nuclei are then resuspended in the low salt lysis buffer that additionally contains sufficient DNAse I and incubated for a time sufficient to efficaciously degrade the DNA in such nuclei to a size and viscosity that allows subsequent purification of the c-Myc from this preparation as described below.

Following the DNAse I treatment, the sample is diluted with a high salt neutral buffer that brings the salt (as NaCl) concentration of the sample to at least 2 M. Such high salt buffer preferably additionally also contains amounts $MgCl_2$ sufficient to maintain the same concentration of this salt in the final diluted preparation, and also additional detergent NP40so as to retain efficacious levels after sample dilution.

In mammalian host cells, c-Myc is generally tightly associated with the nuclei. Accordingly, it is necessary to solubilize c-Myc in a manner that does not destroy its biological activity or its ability to renature into a biologically active form. The residual nuclear material is first removed by

centrifugation and then the pellet resuspended for solubilization of the c-Myc. Solubilization of the c-Myc protein in a manner that destroys this association may be achieved with either sodium dodecyl sulfate (SDS) or urea at concentrations greater than 4 M. Preferably, 5M urea is utilized. Residual non-lysed nuclei may also be solubilized at this time by vigorous stirring for about 30 min. The solution is then centrifuged to pellet any remaining insoluble material prior to the subsequent chromatography steps, for example, at 5000xg for about 10 min.

The supernatant fraction recovered from the centrifugation step is applied to a DEAE Sepharose CL-6B column equilibrated in the urea-containing buffer as described above, and the column thoroughly washed with such buffer to remove unbound protein. A second wash was performed with the addition of an intermediate amount of NaCl, 0.1M NaCl to the buffer. Finally, Myc protein was eluted by raising the salt concentration in the buffer to 0.35M.

All protein eluting with the 0.35M salt wash were collected and applied to a FPLC Mono-Q column. The column was washed and with a gradient of 0.10 M NaCl to 0.35 M NaCl, followed by a 2 M NaCl step wash. Holding the gradient at intermediate salt concentrations, for example at about 0.19 M NaCl, until the end tail of the contaminating protein is finished eluting will enhance the purity of the subsequently eluted Myc protein.

Myc may be identified in the column eluent by any technique that specifically recognizes Myc protein or its activity. For example, a monoclonal antibody such as 1F7 may be used in an immunoassay for the presence of Myc protein. Alternatively, DNA binding activity to an oligonucleotide containing the sequence 5'-CACGTG-3' may be used to monitor the purification. Monoclonal antibody 1F7 is directed against the peptide sequence of amino acids 305-317 in murine c-Myc. Other Myc monoclonal antibodies useful in such assays are commercially available.

Pools of fractions from this column contained the C2 and C2′ binding activities described below, and the presence of peptides capable of entering into C2 and C2′ hetero-oligomers, and especially C2 and C2′ hetero-oligomers, may be assayed by the ability of such hetero-oligomers to bind to the DNA sequences 5'-CACGTG-3' and 5'-CAGCTG-3', respectively. Myc purified from the CHO cells appeared as multiple bands by immunoblot.

b. Purification of c-Myc and Its Partners From a Baculovirus Source

Human c-Myc may also been purified using the baculovirus overexpression system. For purification, Sf9 cells that had been infected with recombinant baculovirus carrying the c-Myc gene, using techniques known in the art were harvested just prior to the onset of lysis (~48 hours post infection). Solubilization and purification of the recombinant c-Myc were carried out as with the CHO produced Myc resulting in a yield of 2.5 mg/8x10^8 cells. Myc purified from these insect cells was apparently homogeneous by silver staining, and ran on electrophoresis as a single diffuse band of ~60kD. This was in contrast to the multiple bands observed with mammalian Myc by immunoblot (Fig. 1B).

c. Detection of Sequence Specific DNA Binding Activity

The above preparations contain two sequence specific DNA-binding activities that both contain Myc protein. The first activity contains only Myc (i.e., forms the Myc homo-oligomer) and binds very weakly to sequences with the core CACGTG. The binding is assayed by determining the off rate and by competitor assays, both techniques known in the art. The binding of c-Myc homo-oligomers is characterized by an immeasurably fast off rate and by the observation that it is almost impossible to add

enough cold competitor sequence to completely compete away this complex in electrophoretic mobility shift assays (EMSA). This latter observation implies that it may not be possible to raise oligonucleotide concentrations above the K_D, thus preventing the determination of exactly what fraction of the final Myc preparations are active for sequence specific binding by the Myc homo-oligomers.

A binding site selection procedure may be used to determine the optimal binding site for Myc. Sites may be selected from a pool of random oligomers, such as 20-mers, in order to decrease bias in determining an optimal binding site. A 12 base consensus sequence of GACCACGTGCTC [SEQ ID No. 1] may be used, with the central E box core of CACGTG appearing to be most conserved. Halazonetis and Kandil (Halazonetis and Kandil, *Proc. Natl. Acad. Sci. USA* 88:6162-6166 (1991)) assumed that the flanking sequences might be symmetric, and reported an optimal sequence of GACCACGTGGTC [SEQ ID No. 2]. This sequence is quite similar to the consensus that is preferred here, differing in only the 10th position (where predominantly a C was utilized in the invention, although G is significantly represented Fig. 7, Group I). Accordingly to the invention, it is possible to select a 12 base consensus sequence from a pool of predicted complexity of 4^{20} ($\sim 10^{12}$) thus indicating that Myc has a strong sequence preference despite its apparent weak binding affinity.

The second Myc containing DNA-binding complex provided in the preparations of the invention also binds to sequences with a core of CACGTG, but binds significantly more tightly than Myc alone. This complex (the C2 complex) requires a 26-29 kD factor in addition to Myc. This additional factor copurified with Myc, presumably because of similar chromatographic properties and not via association with Myc since the chromatography performed in 5M urea would denature such association. This additional factor resembles Max, a protein whose gene was recently isolated from mammalian cells, in that it does not bind efficiently to DNA by itself but can hetero-oligomerize with Myc to bind tightly to the

sequence CACGTG. However, that the factor of the invention differs from Max in its apparent size (Max is reported to migrate at 21 kD). Additionally, the Myc/Max hetero-oligomer appears to migrate at least as slowly as a Myc only complex in EMSAs, while the C2 complex of the invention migrates more rapidly than Myc alone.

In addition to the 26-29 kD factor, a second copurifying factor of 40-50 kD has been identified. The sites selected by complexes containing this factor (herein termed C2′ complexes) contained a CAGCTG core (the μE2 sequence motif) as well as flanking sequences which bear a striking resemblance to a recently reported binding site for myogenin homo-oligomers (Wright *et al.*, *Mol. Cell. Biol. 11*:4104-4110 (1991)). Myogenin is an HLH containing protein of predicted molecular weight 32.5 kD whose optimal binding site is AACAGT/CTGTT [SEQ ID No. 3]. None of the sites (0/36) selected by the C2 or C2′ complexes of the invention contained a CAGTTG motif while roughly half of the myogenin selected sites contained such core sequences.

d. Assay for a Compound that Inhibits Myc Action

For the ease in describing these assays, C1 complex association and/or DNA binding, C2 complex association and/or DNA binding, and C2′ complex association and/or DNA binding are all referred to as c-Myc activity.

Assays for c-Myc activity may be performed *in vitro* or *in vivo*. *In vitro* assays may be performed as described in the Examples, for example, by evaluating the effects the desired compound or various amounts of such compound on the results of the electrophoretic mobility shift assay and site selection techniques that will reveal whether binding of the oligomer of interest to a specific DNA sequence motif has occurred in the presence of the compound.

For the *in vivo* assay of a compound that inhibits the desired Myc activity at least two genetic constructs are utilized. First is required a recombinant construct capable of expressing Myc is required; second is required a reporter gene whose expression is operably linked to the Myc activity and especially to the binding of the desired oligomer to the specific DNA sequence or motif.

If desired, a recombinant construct capable of expressing a C2 complex protein or C2′ complex protein may also be used. Alternatively, a host may be chosen may be chosen that naturally expresses such protein.

Recombinant constructs that are capable of expressing Myc protein may be constructed utilizing the guidelines as described below or purchased commercially.

The desired DNA binding sequence may be operably linked to any gene which confers a selectable marker in the host system. In a preferred embodiment, a marker gene which allows phenotypic selection in yeast, and especially in *Saccharomyces cerevisiae* is used.

Yeast that have been co-transformed with both an expressible *myc* gene and with the desired DNA binding sequence may be used to (1) identify the presence or absence of endogenous host proteins that interact with Myc in a C2 or C2′complex (2) classify a protein as a C1 complex protein or as a C2′ complex protein; and (3) identify and classify compounds as agents which disrupt such Myc activity. C2 complex proteins have previously also been termed Myc "partner" proteins.

All three applications are based on the same principle: in the presence of c-Myc biological activity, one of three things will happen: C1 complexes will form; C2 complexes will form; or, C2′ complexes will form. The protein complexes so formed, and especially the oligomeric complexes, will bind to a specific DNA motif, binding to such motif will be operably linked to the marker gene, and expression of the marker gene will be altered, preferably stimulated, in response to such DNA binding.

In the absence of such oligomerization, oligomer-DNA complex formation will not occur and expression of the marker protein will not be altered.

In the assays of the invention, there may be some level of binding to a desired DNA binding sequence even in the absence of c-Myc. However, when c-Myc is present in the cell, the amount and strength of the specific DNA binding is increased.

Hosts that have been co-transformed with both an expressible c-*Myc* gene and with the desired DNA binding sequence may be used to assay for the presence or absence of endogenous host proteins that interact with c-Myc activity. If such analyses reveal that the host contains c-Myc binding proteins, or c-Myc dependent oligomers which, in the presence of c-Myc specifically bind to a desired DNA sequence, such c-Myc partner protein or dependent-oligomer protein may be isolated using techniques known in the art such as gel mobility shift analysis, cDNA expression cloning vectors such as, for example, λgt10 and λgt11, or other cloning systems specifically designed for high-efficiency cloning and expression of full-length cDNA in yeast such as, for example, pG1 and pTRP56, all of which are commercially available (Clontech, Palo Alto, California).

It is not necessary that the host be completely deficient in C2 complex proteins (c-Myc partner proteins) or C2′ complex proteins to be useful in the method of the invention. As described below, if c-Myc is expressed at levels much greater than those found in the host, reporter gene transcription from endogenous partner proteins may be negligible, or of such low amount that it does not otherwise alter the utility of the methods of the invention.

If the c-Myc expression is transcribed with a strong promoter, and/or if the c-*Myc* expression cassette is supplied on a high copy number vector, the levels of c-Myc will be high enough to overcome a low level background and such c-Myc constructs may be used to analyze the ability of cloned c-Myc partners to influence c-Myc DNA binding. One of

ordinary skill in the art can adapt the expression system to the level of expression desired using methods known in the art.

The C2 complex protein (the partner protein), or the C2´ complex protein, if supplied as a recombinant construct to the host cell, should be capable of expressing at levels comparable to that of the c-Myc protein. C2 complex proteins may be identified by utilizing a phage plaque assay, as described in the commonly-owned, copending U.S. patent application entitled "Protein Partner Screening Assays and Uses Thereof," Application No. 510,254, filed April 19, 1990, and incorporated herein by reference. Proteins identified by such screening assay can be subcloned into eukaryotic expression vectors known in the art and commercially available so as to provide a recombinant source of partner protein gene expression.

The genetic constructs of the invention may be placed on different plasmids, or combined on one plasmid. A construct may also be inserted into the genome of a host cell. Preferably, the construct coding for the c-Myc protein and the construct coding for the C2 complex protein or the C2´ complex protein are provided to the host on two different plasmids.

It is important to establish that the effect of the compound is due to an effect on c-Myc activity and not an effect on the activity of the reporter product *per se*. Such effect can be established by comparing the results found in hosts which lack either the c-Myc expression vector or the C2 or C2´ protein expression vector or both.

The desired DNA binding motif may be located at any site in the transcription cassette of the reporter gene which allows for the transcription of that gene to be operably-linked to binding of the desired oligomer. Thus, such motif may be located 5' to the transcriptional start site or 3' to the transcriptional start site, for example, in an intron, similar to its location relative to the promoter region in the immunoglobulin genes.

The reporter gene whose expression is operably linked to c-Myc activity and especially to oligomer DNA binding may be any gene whose expression can be monitored. Any detectable phenotype change may serve

as the basis for the methods of the invention. In a preferred embodiment, the reporter gene is a gene not normally expressed by the host, or a gene that replaces the host's endogenous gene. Any reporter gene which is capable of being operably-linked to a promoter capable of responding to the binding of the oligomer of interest to the specific target DNA sequence may be used.

Especially, for example, genes that endow the host with an ability to grow on a selective medium are useful. For example, in yeast, use of the yeast LEU2 gene as a reporter gene in strains that normally lack LEU2 allows such yeast to grow on leucine as a sole carbon source. Expression the reporter gene is monitored by merely observing whether the host possesses the ability to grow on leucine. In a similar manner, use of the *suc2* gene as a reporter gene would allow growth of the a $suc2^-$ yeast host on sucrose to be used as the detection method. In both examples, growth on the indicated substrate could be used to indicate specific DNA binding of the oligomer of interest and lack of such growth could be used to indicate lack of binding or lack of oligomer formation.

In another example, a construct (and host) which is $gal1^+gal10^-$ would respond to galactose in the medium; a construct (and host) which is $lac2^+gal1^+$ would be lactose sensitive. Other reporter genes include *his3*, *ura3* and *trp5*. One of ordinary skill in the art can imagine many other appropriate reporter systems which would reveal the presence or inhibition of DNa binding or biological activity of the oligomer of interest.

Reporter constructs in which the desired DNA sequence motif and the *lacZ* reporter gene are operably linked will express β-galactosidase in response to binding of a c-Myc activity induced oligomer binding to such DNA sequence. Such expression can be easily scored by monitoring the ability of the host to produce β-galactosidase (Maniatis, T. *et al.*, *Molecular Cloning (A Laboratory Manual)*, 2nd edition, Cold Spring Harbor Laboratory, 1989). The production of β-galactosidase may be visually monitored by detecting its activity to reduce the chromophoric dye,

X-gal (commercially available from International Biotechnologies, Inc., New Haven, CT). β-galactosidase reduces X-gal to a form which possesses a blue color. In another embodiment, the coding sequence of chloramphenicol acetyltransferase (CAT) is used as the reporter gene.

Any detection method that can identify expression of the reporter gene may be used. For example, levels of the product of the reporter gene may be directly assayed with an immunoassay. Such immunoassays include those wherein the antibody is in a liquid phase or bound to a solid phase carrier. In addition, the reporter gene can be detectably labeled in various ways for use in immunoassays. The preferred immunoassays for detecting a reporter protein using the include radioimmunoassays, enzyme-linked immunosorbent assays (ELISA), or other assays known in the art, such as immunofluorescent assays, chemiluminescent assays, or bioluminescent assays.

In an assay to screen for the ability of a compound to alter binding of the oligomer of interest, yeast strains that express such the desired peptide or peptides and which contain the related DNA binding sequence motif, may be plated and grown as lawns and the compound to be tested may be applied to the plates on a filter paper disk that is impregnated with such compound. Alternatively, the compound may be incorporated into the media within which the host cells are growing.

One may be able to detect the ability of a compound to alter c-Myc activity by the appearance of a zone, which often resembles a halo, around the compound-impregnated disk. If for example, the compound is toxic to the host's survival *per se*, the host will not grow in the zone containing the compound.

The methods of the invention can be used to screen compounds in their pure form, at a variety of concentrations, and also in their impure form. The methods of the invention can also be used to identify the presence of such inhibitors in crude extracts, and to follow the purification of the inhibitors therefrom. The methods of the invention are also useful in

the evaluation of the stability of the inhibitors identified as above, to evaluate the efficacy of various preparations.

The permeability of cells to various compounds can be enhanced, if necessary, by use of a mutant cell strain which possess an enhanced permeability or by using compounds which are known to increase permeability. For example, in yeast compounds such as polymyxin B nonapeptide may be used to increase the yeast's permeability to small organic compounds. In cells from the higher eukaryotes, dimethyl sulfoxide (DMSO) may be used to increase permeability. Analogs of such compounds which are more permeable across yeast membranes may also be used. For example, dibutyryl derivatives often display an enhanced permeability.

In a preferred embodiment, the genetic constructs and the methods for using them are utilized in eukaryotic hosts, and especially in yeast, insect and mammalian cells. The introduced sequence is incorporated into a plasmid or vector capable of either autonomous replication or integrative activity.

The sequence of c-Myc is known (Battey, J. *et al.*, *Cell 34*:779-787 (1983)) and probes which are capable of identifying a c-*Myc* clone are commercially available (New England Nuclear/DuPont Biotechnology Boston, MA).

The DNA sequence of the desired gene may be chemically constructed if it is not desired to utilize a clone of the genome or mRNA as the source of the genetic information. Methods of chemically synthesizing DNA are well known in the art (*Oligonucleotide Synthesis, A Practical Approach*, M.J. Gail, ed., IRL Press, Washington, D.C., 1094; *Synthesis and Applications of DNA and RNA*, S.A. Narang, ed., Academic Press, San Diego, CA, 1987). Because the genetic code is degenerate, more than one codon may be used to construct the DNA sequence encoding a particular amino acid (Watson, J.D., In: *Molecular Biology of the Gene*, 3rd edition, W.A. Benjamin, Inc., Menlo Park, CA, 1977, pp. 356-357).

To express the recombinant constructs of the invention, transcriptional and translational signals recognizable by the host are necessary. A cloned protein encoding DNA sequence, obtained through the methods described above, (preferably in a double-stranded form), may be operably-linked to sequences controlling transcriptional expression in an expression vector, and introduced, for example by transformation, into a host cell to produce recombinant proteins useful in the methods of the invention, or functional derivatives thereof. Such techniques are well known in the art (*Recombinant DNA Methodology*, Wu, R. *et al.*, eds., Academic Press, (1989); Maniatis, T. *et al.*, *Molecular Cloning (A Laboratory Manual)*, second edition, Cold Spring Harbor Laboratory, 1989).

Transcriptional initiation regulatory signals can be selected which allow for repression or activation of the expression of the c-Myc construct or construct of the recombinant C2 complex peptide (or the C2′ peptide), or both, so that expression of such constructs can be modulated, if desired. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical regulation, for example, by a metabolite, salt, or substrate added to the growth medium.

Where the native expression control sequences signals do not function satisfactorily in the host cell, then sequences functional in the host cell may be substituted.

Expression of the constructs of the invention in different hosts may result in different post-translational modifications which may alter the properties of the proteins expressed by these constructs. It is necessary to express the proteins in a host wherein the ability of the protein to retain its biological function is not hindered. Expression of proteins in yeast hosts is preferably achieved using yeast regulatory signals. The vectors of the invention may contain operably-linked regulatory elements such as

upstream activator sequences in yeast, or DNA elements which confer species, tissue or cell-type specific expression on an operably-linked gene.

In general, expression vectors containing transcriptional regulatory sequences, such as promoter sequences and transcription termination sequences, are used in connection with a host. These sequences facilitate the efficient transcription of the gene fragment operably-linked to them. In addition, expression vectors also typically contain discrete DNA elements such as, for example, (a) an origin of replication which allows for autonomous replication of the vector, or, elements which promote insertion of the vector into the host's chromosome in a stable manner, and (b) specific genes which are capable of providing phenotypic selection in transformed cells. Eukaryotic expression vectors may also contain elements which allow it to be maintained in prokaryotic hosts; such vector are known as shuttle vectors.

The precise nature of the regulatory regions needed for gene expression will vary between species or cell types and there are many appropriate expression vector systems that are commercially available.

In a highly preferred embodiment, yeast are used as the host cells. The elements necessary for transcriptional expression of a gene in yeast have been recently reviewed (Struhl, K. *Ann. Rev. Biochem.* 58:1051-1077 (1989)). In yeast, most promoters contain three basic DNA elements: (1) an upstream activator sequence (UAS); (2) a TATA element; and, (3) an initiation (I) element. Some promoters also contain operator elements. Methods in yeast genetics are well known (Struhl, K. *Nature 305*:391-397 (1983); Sherman, *et al., Methods in Yeast Genetics,* Cold Spring Harbor Laboratory (1983)).

In another embodiment, mammalian cells are used as the host cells. A wide variety of transcriptional and translational regulatory signals can be derived for expression of proteins in mammalian cells and especially from the genomic sequences of viruses which infect eukaryotic cells.

Once the vector or DNA sequence containing the construct(s) is prepared for expression, the DNA construct(s) is introduced into an appropriate host cell by any of a variety of suitable means, for example by transformation. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene sequence(s) results in the production of the protein. This expression can take place in a continuous manner in the transformed cells, or in a controlled manner.

Genetically stable transformants may be constructed with episomal vector systems, or with integrated vector systems whereby the fusion protein DNA is integrated into the host chromosome. Such integration may occur *de novo* within the cell or be assisted by transformation with a vector which functionally inserts itself into the host chromosome, for example, with retroviral vectors, transposons or other DNA elements which promote integration of DNA sequences in chromosomes.

Cells which have been transformed with the DNA vectors of the invention are selected by also introducing one or more markers which allow for selection of host cells which contain the vector, for example, the marker may provide biocide resistance, e.g., resistance to antibiotics, or heavy metals, such as copper, or the like.

The transformed host cell can be fermented or cultured according to means known in the art to achieve optimal cell growth, and also to achieve optimal expression of the cloned protein sequence fragments. As described hereinbelow, a high level of recombinant protein expression for the cloned sequences coding for the proteins can be achieved according to a preferred procedure of this invention.

The methods of the invention are not intended to be limited to c-Myc and possess utility for the characterization of inhibitors against any Myc protein, such as, for example, N-Myc and L-Myc. The C2 complex peptides of the invention may interact with more than one Myc protein and

the C2′ complex peptides of the inventions may form as the result of the activity of more than one Myc protein.

The following examples further describe the materials and methods used in carrying out the invention. The examples are not intended to limit the invention in any manner.

EXAMPLES

Example 1

Materials and Methods

Cell Growth and Myc Overexpression: The 5A cell line was maintained in spinner culture under selection with 80 μM methotrexate. Protein purification started with roughly 6 liters of cells at $8x10^5$/ml grown up without selection. Heat shock promoter induction was achieved by resuspension in preheated fresh media (43°C) at 1/3 the original volume. Cells were incubated with stirring at 43°C for 1 h. To allow translation of the accumulated mRNA, cells were transferred to 37°C culture conditions for 3 h. Cells were then subjected to the purification described below.

The baculovirus overexpression vector was constructed by insertion of the BamH1/BclI fragment of pGEMMycB [Halazonetis and Kandil, *Proc. Natl. Acad. Sci. USA* 88:6162-6166 (1991)] into the BamH1 site of a baculovirus expression vector, pVL941, obtained from the laboratory of Max Summers (Texas A&M University, College Station, Texas). The resulting plasmid contained the entire coding sequence of the human Myc gene including 6 nucleotides 5′ of the initiation codon and 3′ untranslated sequence extending to the genomic RsaI site. Sf9 cells were grown and infected with recombinant baculovirus according to the methods of Summers [Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experiment Station Bulletin No. 1555] with minor changes. Cells were passaged in

spinner culture and plated on 150 mm diameter tissue culture plates for protein production. Cells were infected and harvested approximately 48 h post infection by scraping. Cells were then washed in PBS and subjected to the purification described below.

The Protein A-c-Myc fusion protein was expressed in the *E. coli* AR68 strain from a previously published pRIT2T vector [Dang, C.V., *Anal. Biochem. 174*:313-317 (1988)] which fused the Ig binding portion of protein A to either amino acids 353-439 or amino acids 372-439 of c-Myc. Growth and induction of the cells was as per Dang *et al.* [*Anal. Biochem. 174*:313-317 (1988)].

Protein Purification: All purification steps were carried out on ice or with ice cold buffers unless otherwise stated. Cells may be used fresh or stored quick frozen in liquid nitrogen for larger batch preparations. 5A or Sf9 cells were washed in phosphate-buffered saline (PBS) and resuspended at $2.1x10^7$ cells/ml in Low Salt Lysis Buffer (20 mM HEPES pH 6.8, 5 mM KCl, 5 mM $MgCl_2$, 0.5% NP40, 0.1 % Na-deoxycholate, 1 μg/ml aprotinin, and 0.1 mM PMSF) [Evan and Hancock, *Cell 43*:253-261 (1985)]. After 10 min cells were subjected to 40 strokes in a Dounce homogenizer with a type A pestle. Nuclei were pelleted at 1000xg, 5 min, 4°C, washed once in 50 ml Low Salt Lysis Buffer, resuspended at $2.5x10^8$ nuclei/ml in Low Salt Lysis Buffer containing 50 μg/ml DNAse I and incubated at 4°C for 1 h. An equal volume of ice cold 2X High Salt Buffer (2x concentrations: 20 mM Tris, pH 7.4, 4 M NaCl, 1 mM $MgCl_2$, and 0.1% NP40) [Evan and Hancock, *Cell 43*:253-261 (1985)] was then added, mixed gently, and incubated for 10 min. The residual nuclear material (including the c-Myc protein) was pelleted (2000xg, 10 min, 4°C) and resuspended for solubilization at $5.5x10^7$ nucleus equivalents/ml in Buffer A (50 mM Tris, pH 8.0, 2 mM EDTA, 5 % glycerol, .1 mM DTT, and .1 mM PMSF) [Watt *et al.*, *Mol. Cell. Biol. 5*:448-456 (1985)] containing 5 M urea (referred to as 5 M urea Buffer A) achieved by

dilution of a freshly deionized stock of 6 M urea. This and all buffers used on columns were passed through 0.2 pore μm filter units. Residual nuclei were solubilized by vigorous stirring on ice for 30 min. This protein solution was centrifuged (10 min, 5000xg, 4.C) to pellet any insoluble material prior to chromatography. The supernatant was loaded on a 10 ml DEAE Sepharose CL-6B (Pharmacia) column equilibrated with 5 column volumes of 5 M urea Buffer A. Sample loading was at 0.1 ml/min and column washing and elution were at 0.4 ml/min. After loading, the column was washed with 3 volumes 5 M urea Buffer A containing no additional salt followed by 4 volumes of the same buffer containing 0.1 M NaCl. Myc protein was eluted in the following elution step at 0.35 M NaCl. The protein containing fractions of this 0.35 M NaCl step were pooled and diluted with fresh 5 M urea Buffer A to 0.1 M NaCl and loaded onto a 1 ml FPLC Mono-Q column (Pharmacia) run at 0.5 ml/min. The Mono-Q column was eluted with a programmed gradient of 5 ml spanning 0.10 M NaCl to 0.35 M NaCl followed by a 2 M NaCl step. For enhanced purity the gradient was held manually at approximately 0.19 M until the major contaminating protein finished eluting as determined by an in line UV monitor. In the initial development of the purification protocol fractions from the columns were assayed for Myc by slot blotting followed by visualization using the 1F7 monoclonal antibody and ^{125}I-labeled secondary antibody. For later preparations silver staining of SDS-PAGE allowed sufficiently unambiguous identification of the Myc proteins and provided an assessment of the purity of given fractions. The Myc containing fractions were pooled based on purity and dialyzed against buffer containing 20 mM Tris, pH 7.8, 50 mM KCl, 10 % glycerol, 0.1 mM DTT, and 0.1 mM PMSF (referred to as Dialysis Buffer) in bags of SpectroPor 2 membrane for 3 changes, 2 liters each, for a minimum of 3 h each. Pools of fractions prepared this way contained C1 and C2 (and C2') binding activities. To obtain pure C1 binding activity the Myc-containing

Mono Q fractions were assayed by EMSA and those free of C2 binding activity were pooled and dialyzed separately.

The bacterially produced Protein A-c-Myc fusion protein was partially purified by differential centrifugation and solubilized in 5 M urea according to Watt *et al.* [Bagchi *et al.*, *Mol. Cell. Biol. 7*:4151-4158 (1987)] with the following minor modifications: Protease inhibitors were present in the initial lysis buffer (10 μg/ml pepstatin, 1 mM PMSF, 50 μg/ml aprotinin, 2 μg/ml leupeptin, 10 mM Na-metabisulfite, and 1 mM benzamidine) and cells were sheared by 6 bursts of 15 s each in a Cuisinart MiniMate on ice. The urea solubilized material was cleared of insoluble material by centrifugation (10,000xg, 10 min, 4°C) and dialyzed into Dialysis Buffer containing 0.5 mM DTT. Precipitated material was removed by centrifugation (15,000xg, 20 min, 4°C). Protein A-Myc fusion protein was purified from the supernatant by IgG affinity essentially according to Nilsson *et al.* [*EMBO J. 4*:1075-1080 (1985)]. A 1 ml aliquot of supernatant was incubated with 0.1 ml of a 50% slurry of IgG Sepharose 6 fast flow (Pharmacia) rocking for 1 h at 4°C. The pellet was washed twice with Buffer A and the fusion protein eluted with 0.3 M lithium diiodosalicylate (LIS). The eluate was then dialyzed extensively to remove the LIS (initially against Buffer A at room temperature to avoid LIS precipitation, then against Dialysis Buffer 4°C). The two bacterially expressed Myc preparations were compared by Coomassie staining of SDS-PAGE to ensure that equal amounts of the fusion proteins were used for experiments.

N-Terminal Sequencing: The 3 bands of purified Myc from 5A cells were individually isolated by electroelution according to Hunkapiller *et al.* [*Meth. Enz. 91*:227-236 (1983)]. Preparative SDS-PAGE was carried out and protein bands excised after visualization with Coomassie Brilliant Blue R-250. Alter electroelution the material was precipitated 2 times with

methanol/acetone and submitted for N-terminal sequencing by Edman degradation.

Antibodies: The monoclonal antibody, 1F7 (a generous gift of R. Chizzonite, Hoffman LaRoche), is directed against the peptide sequence comprising amino acids 305-317 in murine c-Myc [Miyamoto *et al.*, *Proc. Natl. Acad. Sci. USA* 82:7232-7236 (1985)]. The antibody directed against cI was monoclonal 51F [Breyer and Sauer, *J. Biol. Chem.* 264:13348-13354 (1989)] which had been purified by ammonium sulfate precipitation and chromatography on QAE Sephadex.

Electrophoretic Mobility Shift Assay (EMSA): Radiolabeled probes were produced via a Klenow fill in of annealed oligonucleotides containing 4 base 5' overhangs at each end (see table below for sequences). Binding reactions took place in a final volume of 20 μl containing 2 ng of labeled probe, 125 ng poly d(IC), an indicated amount of protein, and the following final buffer conditions: 10 mM Tris, pH 7.5, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 1 mM $MgCl_2$ and 5% glycerol. Binding reactions were allowed to proceed for 20 min at room temperature and were then loaded immediately on a 4% polyacrylamide gel which had been prerun at least 1 h at 10V/cm. Electrophoresis was for 1.5 h at 10V/cm in 0.5x TBE.

Cut and Renature: The method of Bagchi *et al.* [Bagchi *et al.*, *Mol. Cell. Biol.* 7:4151-4158 (1987)] was followed except for the final dialysis step. Precipitated protein samples containing BSA as carrier protein were solubilized in 6 M guanidine-hydrochloride (200 μl unless otherwise indicated) according to Bagchi *et al.* [*Mol. Cell. Biol.* 7:4151-4158 (1987)]. Directly prior to analysis by EMSA the samples were subjected to dialysis alone or in combination with another sample in a total volume of 15 μl.

Equal volumes of each sample were used in a given experiment and the volume was brought to 15 μl using 6 M GuHCl containing 0.1 mg/ml BSA. Dialysis was against 40 ml of Dialysis Buffer carried out for 1 h at 4°C on floating 13 mm membrane discs (Millipore #VSWP-013, pore size 0.025 μm).

Site Selection from Random Sequences: The following procedure was devised based on the method of Pollock and Treisman [*Nucl. Acids Res.* *18*:6197-6204 (1990)]. A 52 base oligonucleotide "randomer" (see table below) was annealed to the following 16 base primer: Xho I primer 5' CCGATATCTCGAGACGG 3', [SEQ ID No. 4]. The annealed primer was extended using Klenow and nucleotides (0.2 mM cold dNTPs and 0.4 μM α^{32}P-dCTP 800Ci/mmol) to create a pool of double stranded probes representing approximately 4^{20} sequences. The initial round of binding site selection by EMSA utilized 200 ng of this pool and either 0.37 μg of baculovirus produced c-Myc or 0.5 μg of CHO produced c-Myc. Other parameters were as previously described for EMSA. Lanes containing randomer probes were alternated with reference lanes containing 2 ng $(USE)_3$ probe and 0.37 μg of baculovirus c-Myc. The completed EMSA gel was electroblotted onto NA45 membrane (200 mA, 2.5 hrs) and the wet membrane was wrapped in plastic wrap and exposed for at least 1.5 hrs. The regions of the randomer lanes corresponding to the visible C1 and C2 complexes of the reference lanes were excised and eluted with 100 of elution solution (10 mM Tris, pH 8.0, 1 mM EDTA, 1 M NaCl) 30 min at 68°C. The liquid was transferred to a fresh tube and the membrane was rinsed with 100 μl TE which was added to this eluate. After pelleting the particulate debris, the DNA was precipitated with the addition of 10 μg glycogen, 2 μl 1 M $MgCl_2$ and 2.5 volumes of ethanol. The pellet was rinsed with 70% ethanol, dried, and the recovery assessed by scintillation counter. The entire pellet of each sample (~29-57 pg) was resuspended in 10 μl 10x PCR buffer (500 mM KCl, 100 mM Tris, pH 8.4, 1 mg/ml

gelatin, 15 mM $MgCl_2$) and 32 μl water. After addition of 1 μl each of 100 μM Xho I primer and Xba I primer (5' GGACGATCTAGATTCG 3', [SEQ ID No. 5]), 5 μl of nucleotide mix (2 mM dNTPs and 4 μM α^{32}P-dCTP 800Ci/mmol), and 1 U Taq polymerase the reactions were overlaid with paraffin oil and subjected to 20 cycles of PCR in an Ericomp machine: 2 min 94°C, 20x (15 sec 95°C, 15 sec 55°C), 10 min 72°C. The products were gel purified on 10% acrylamide and precipitated using 10 μg glycogen as carrier. Recovery was measured by scintillation counter and after resuspension in the EMSA reaction buffer (10 mM Tris, pH 7.5, 50 mM KCl, 1 mM EDTA, 1 mM $MgCl_2$, and 5% glycerol) this probe was used for the next round of EMSA selection. Subsequent cycles were primarily as above, however, 50 ng of probe was used. Eight rounds of selection and amplification were completed for the baculovirus c-Myc and seven rounds for the CHO c-Myc. After the final PCR reaction the products were extracted twice with phenol, twice with ether, and precipitated prior to digestion with Xho I and Xba I. After gel isolation the appropriate fragment was subcloned into the Bluescript SK vector (Stratagene) and sequenced by standard procedures.

Oligonucleotides Used: Oligonucleotide sequences that were used are shown below, with the E-Box core sequences underlined:

SEQ ID NO. 6:
$(\mu E2)_3$ 5' GATCTCTGCAGCAGCTGGCAGCAGCTGGCAGCAGCTGGCG 3';

SEQ ID NO: 7:
$(\mu E3)_3$ 5' GATCTGCAGTCATGTGGCGTCATGTGGCGTCATGTGGCAG 3';

SEQ ID NO: 8:
$(USE)_3$ 5' GATCTGCAGTCACGTGGCGTCACGTGGCGTCACGTGGCAG 3';

SEQ ID NO. 9:

MLC-A 5′ TCGACGTCGCAGCAGGTGCAG 3′;

SEQ ID NO. 10:
MLC-B 5′ TCGACCCCACCAGCTGGCGAG 3′;

SEQ ID NO. 11:
ERP1/2 5′ AGCTTCGAACACCTGCAGCAGCTGGCAGGAAGCAGGCCTA 3′;

SEQ ID NO. 12:
ERP3/4 5′ AGCTTTAAAATCCCCACCAGCTGGCGAAGCAACAGGTGCA 3′;

SEQ ID NO. 13:
HSE 5′ AATTGCGAAACCCCTGGAATATTCCGACCTGGCAGCCTC 3′;

SEQ ID NO. 14:
SMS 5′ TCGACTTTAGACCACGTGGTCCCCTCGA 3′;

SEQ ID NO. 15:
Randomer 5′ GGACGATCTAGATTCG$(N)_{20}$CCGTCTCGAGTATCGG 3′.

Example 2

Purification of c-Myc Protein

A primary goal of this work was to purify and characterize Myc from a mammalian source. An inducible mammalian overexpression system that has been described previously was utilized (Wurm *et al.*, *Proc. Natl. Acad. Sci. USA 83*:5414-5418 (1986)). Briefly, the two coding exons of the mouse c-Myc gene under the control of a *Drosophila* heat shock promoter had been integrated and amplified in the genome of a Chinese hamster ovary (CHO) cell line. This overexpressing cell line, 5A, was adapted to spinner culture. Heat shock (43°C) induces transcription of the amplified *myc* genes while a subsequent 2 hour recovery period at normal growth temperature (37°C) permits translation. The resulting products were

phosphoproteins of 60, 62, and 72kD which were immunoprecipitable with Myc-specific monoclonal antibodies (Wurm *et al.*, *Proc. Natl. Acad. Sci. USA* 83:5414-5418 (1986)). The c-Myc produced was tightly associated with the nuclei and attempts to solubilize it using a number of detergents, salts, and reducing agents were unsuccessful (data not shown). Significant solubilization was achieved however with either SDS or with urea at concentrations greater than 4 M. For purification, the Myc was solubilized with 5 M urea and chromatographed on DEAE resin and FPLC Mono-Q as described in materials and methods. The presence of Myc in the column fractions was assayed by immunoblot using an antipeptide monoclonal antibody, 1F7 (Miyamoto *et al.*, *Proc. Natl. Acad. Sci. USA* 82:7232-7236 (1985)). This purification procedure yielded 150μg of c-Myc per liter of spinner cells ($8x10^8$ cells). The Myc appeared to be 95% homogeneous as judged by silver staining (Fig. 1A).

An alternative translation start site for c-Myc accounts for some of the molecular weight heterogeneity of c-Myc translated *in vitro* and expressed in several cell lines (Hann *et al.*, *Cell* 52:185-195 (1988)). This alternate site is upstream from the canonical start site, however, and is not present in our overexpressor gene. N-terminal sequence analysis of each of the three prominent Myc bands described above revealed, as expected, the sequence predicted by the canonical start site (data not shown), although the N terminal methionine was not present, presumably because of N terminal processing. Therefore the potentially important differences in apparent molecular weight that are observed might be attributed to post-translational modifications and not N-terminal heterogeneity.

Human c-Myc has also been purified using the baculovirus overexpression system. For purification, Sf9 cells that had been infected with recombinant virus were harvested just prior to the onset of lysis (~48 hours post infection). Myc produced using the baculovirus system has been previously reported to be both phosphorylated and tightly associated with the nucleus (Miyamoto *et al.*, *Mol. Cell. Biol.* 5:2860-2865 (1985)).

Solubilization and purification were carried out as with the CHO produced Myc resulting in a yield of 2.5 mg/$8x10^8$ cells. Myc purified from these insect cells was apparently homogeneous by silver staining, and ran as a single diffuse band of ~60kD (Fig. 1B). This was in contrast to the multiple bands observed with mammalian Myc by immunoblot (Fig. 1B).

Discussion

Myc was purified to near homogeneity from overexpressing mammalian cells and baculovirus infected cells. The mammalian derived protein appears to be highly modified in contrast to Myc expressed in and purified from insect cells. Up to 19 distinct species of c-Myc can be identified by two dimensional gel electrophoresis (Fig. 1). These species differ both in size (approximate MRs of 60,000, 62,000 and 72,000, although this estimate of size can vary with different gel conditions) and in pI. These differences in pI might in part be attributed to differences in phosphorylation, as c-Myc is known to be phosphorylated and the change in pI of the species is consistent with incremental additions of phosphate. Although the Myc produced by the baculovirus overexpression system does not demonstrate the same molecular weight heterogeneity as the mammalian protein, it too is phosphorylated (Miyamoto *et al.*, *Mol. Cell. Biol.* 5:2860-2865 (1985)). The specific sites of phosphorylation have not been determined for either Myc preparation and other as yet unidentified modifications may distinguish these two Myc preparations.

Example 3

Specific DNA Binding Activity Present in Purified c-Myc

The presence of a B-HLH domain in c-Myc suggested that it would bind to an E-Box-like sequence of the general pattern CANNTG. These sites were first identified in immunoglobulin enhancers but have since been

found in many other tissue specific enhancers. It was first determined if any of these would be bound by the purified c-Myc proteins described in Example 2. A large number of E box related sequences were screened by electrophoretic mobility shift assays (EMSA). Those shown in Fig. 2 include synthetic oligonucleotides containing trimers of either the μE2 (CAGCTG) or μE3 (CATGTG) sites of the immunoglobulin enhancer and a trimer of the Adenovirus major late promoter upstream element (USE) (CACGTG). Two sites from the myosin light chain (MLC) enhancer are also shown: the A site (CAGGTG) which resembles the kE2 immunoglobulin enhancer site, and the B site (CAGCTG) which has the same core sequence as the μE2 site. The heat shock element (HSE) served as a control since its sequence does not resemble an E-Box core.

Three specific binding activities were detected in this assay forming complexes referred to as C1 (USE specific), C2 (USE specific), and C2′ (μE2 specific). As demonstrated below, despite the comigration of C2 and C2′, these represent separate complexes based on observed differences in protein composition as well as binding specificity. The data presented argue that the C1 complexes are formed by homo-oligomers of Myc while formation of the C2 and C2′ complexes each require an additional protein. The slowly migrating complex (C1) formed most readily on the USE (Fig. 2, lanes 5, 11, and 12), less well on the similar μE3 site (Fig. 2, lanes 2 and 8), and not at all on the other E-Box and non-E-Box sites tested. CHO and baculovirus Myc preparations were similar with regard to the C1 complex, however they differed with regard to the faster migrating complexes. In the mammalian Myc assays the C2′ complex formed on the μE2 site of the immunoglobulin enhancer and the is μE2-like sequence of the MLC-B site (Fig. 2, lanes 1 and 4). Baculovirus Myc contained no binding activity with this specificity (Fig. 2, lanes 7 and 10). In contrast, formation of the C2 complex was detected using either Myc preparation. The C2 complex formed most readily on the USE site (Fig. 2, lanes 5, 11, and 12) and less well on the similar μE3 sequence (Fig. 2, lanes 2 and 8).

Very little if any binding was detected on the kE2-like sequences (MLC-A Fig. 2, lanes 3 and 9, and μE5, data not shown). No specific binding was found on non-E-Box sequences such as the HSE (Fig.2, lane 6 and 13).

Competition experiments were performed on the three binding activities C1, C2, and C2′ to further characterize their specificity (data not shown). In experiments using μE2, μE3, USE, μE5, or HSE sequences as competitors, competition of the C2′ complex formed on the μE2 probes was most easily achieved with the μE2 oligos while the C2 complexes were preferentially competed by the USE sequence. The C1 complex was also competed most efficiently by the USE sequence. A detailed analysis of the binding specificities of these complexes is presented below.

Example 4

Proteins Responsible for Formation of C1, C2, and C2′ Complexes

One scenario suggested by the differences in binding is that Myc might not be the only protein involved in formation of the three complexes. To distinguish the role of c-Myc from other copurifying proteins in the formation of the observed complexes cut and renature experiments were performed as follows. Preparative amounts of Myc were separated by SDS-PAGE. Proteins were electroeluted from various molecular weight slices, precipitated, solubilized in guanidine-hydrochloride and dialyzed to renature for analysis by EMSA. The C1 complex binding activity may be renatured from the Myc containing slices of either baculovirus or mammalian preparations (Fig. 3) while no other slices from the entire gel contained C1 activity (data not shown). These data argue that Myc alone is the protein responsible for the C1 complex, and that full length Myc protein as expressed in eukaryotic cells can bind specifically to sites with the core sequence CACGTG.

Analysis of the proteins responsible for formation of the C2 and C2′ complexes was achieved with additional cut and renature experiments

performed as described above. EMSA using the USE probe revealed no single slice from CHO or baculovirus preparations which contained detectable C2 binding activity (data not shown). However, this activity was recovered by renaturing proteins from a 26-29 kD slice together with proteins in the 60-70 kD Myc containing slice (Fig. 4, lanes 1-8). The 26-29 kD component was present in gels loaded with either CHO or baculovirus produced c-Myc, and, when renatured with Myc, demonstrated the same specificity as the C2 complex in the loaded material. Renaturation of the 26-29 kD slice with BSA or protein A did not yield USE binding activity suggesting that Myc plays a specific role in the recovery of C2 binding activity.

To examine further the roles of copurifying proteins and of Myc modifications in the observed binding, Myc was also purified from a bacterial overexpression system. The expression system and purification method used were those of Chi Dang and colleagues (see materials and methods). The bacterially produced protein contains the IgG binding segment of protein A fused to the C-terminal 85 amino acids of Myc, the segment of Myc which contains the B-HLH and leucine zipper motifs. For many of the B-HLH proteins, the small region of the protein containing the B-HLH motif is not only necessary but fully sufficient for DNA binding if the correct oligomerization partner is present. This protein was able to form the C1 complex on the USE probes (Fig. 4, lane 9) and to combine with the 26-29 kD factor to create the C2 complex (Fig. 4, lane 10). Competition experiments confirmed the specificity of this reconstituted C2 complex. The C1 and C2 complexes formed using this bacterial fusion protein migrated more rapidly than those formed using full length c-Myc (compare Fig. 4, lane 8 with lanes 9 and 10). This may be due to the difference in size between the full length c-Myc (60-72 kD) and the protein A-Myc fusion protein (~38 kD) and therefore the mobility of the C2 complex may be interpreted as an indication that Myc is physically present

in the C2 complex, presumably as part of a hetero-oligomer with the 26-29 kD factor.

Analogous experiments were carried out using a similar bacterial fusion protein containing only the C-terminal 67 amino acids of c-Myc. This protein contains most of the HLH domain and the entire leucine zipper domain but no basic region. Although this protein is capable of forming homo-oligomers in solution (Gentz *et al.*, *Science 243*:1695-1699 (1989)), it was unable to bind to DNA to form the C1 complex and was also unable to combine with the 26-29 kD factor to create any USE binding activity (Fig. 4, lane 12). These data argue that the role of Myc in the C2 hetero-oligomer requires an intact basic region, the region responsible for specific DNA contacts in other B-HLH proteins.

Using cut and renature experiments the μE2 binding activity responsible for the C2′ complex was able to be identified. A small amount of the C2′ complex was frequently seen with proteins from the slice encompassing the 40-50 kD molecular weight range of mammalian Myc preparations (Fig. 5A). Although no C2′ complex was ever seen with the Myc containing slice alone, renaturation of the protein from the Myc slice with the 40-50 kD slice reproducibly increased the amount of C2′ complex formed. Both the baculovirus produced Myc and the bacterially expressed fusion protein containing the basic region, which do not form complexes themselves on μE2 probes, were also able to increase the amount of complex formed by the 40-50 kD slice obtained from mammalian preparations (Fig. 5B and C). Surprisingly the bacterially produced Myc lacking the basic region could also reconstitute C2′ activity, while various other proteins tried including BSA, immunoglobulins, and protein A could not. The apparent lack of a role for this basic region suggests that Myc's involvement in formation of this complex may be other than contacting DNA.

To further determine whether Myc was present in the analyzed complexes, the Myc preparations were incubated with a Myc-specific monoclonal antibody prior to EMSA. The probe used in this experiment

(SMS) contained a single site with the USE core sequence, CACGTG. The Myc-specific antibody eliminated both the C1 and C2 complexes and produced a prominent complex of slower mobility (Fig. 6). It is not clear from these data which of the two complexes was supershifted but the presence of one predominant shifted complex when antibody is present and two complexes in the absence of antibody argues that the Myc-specific antibody also completely disrupted one of the original complexes. There was no effect of a control monoclonal antibody on the formation of either the C1 or C2 complex. The Myc-specific antibody did not alter the C2′ complex, suggesting that Myc is not present in this complex.

From these experiments it can be concluded that the C1 complex is formed by Myc alone, that the C2 complex contains Myc and a 26-29 kd factor and that the C2′ complex contains a 40-50 kd factor but does not contain Myc. It is intriguing that the C2′ complex requires the presence of Myc for formation, but apparently does not contain Myc. Myc therefore appears capable of affecting the 40-50 kd factor's ability to form the C2′ complex without being a member of the complex. Whatever the mechanism, the increase in μE2 binding activity of the 40-50 kD factor appears to be Myc-specific since four different Myc proteins increased the amount of C2′ complex observed while several other proteins did not.

Max protein can be immunoprecipitated from avian and human cells and low stringency Southern analysis has suggested that a single Max gene or a small family of genes exist in other vertebrates as well (Blackwood and Eisenmann, *Science 251*:1211-1217 (1991)). It is possible that hamster and insect cells have an equivalent of Max. The recovery of a Max-like activity from insect cells is particularly interesting since no Myc homologs have been found in insects to date. Drosophila clearly uses B-HLH heterodimers to regulate aspects of development and the possibility remains that the natural partner for the 26-29 kD protein in insect cells is an as yet unidentified B-HLH protein which functions like Myc.

The presence of the 26-29 kD factor in these preparations might limit their usefulness for certain experiments. By pooling Myc containing fractions based on an EMSA assay, one may obtain fractions that contain only the C1 activity and that do not contain the C2 activity, although this modification reduces the final yield by approximately 80%.

Example 5

Selection of Binding Sites For Myc From Random DNA Sequences

In order to determine the optimal binding sites for the three complexes in the Myc preparations described above, a modification of a recently described technique for isolating preferred binding sites from large pools of randomized DNA sequences was used (Pollock and Treisman, *Nucl. Acids Res. 18*:6197-6204 (1990)). Briefly, a pool of double stranded oligonucleotides was created that consisted of 16 base flanking regions of defined sequence surrounding a 20 base region of completely random sequence. Each of the eukaryotic Myc preparations described above was mixed with this pool of sequences and the protein DNA complexes that formed were separated by EMSA. The DNA that ran at the position of the C1 or C2 (and comigrating C2′) complexes was isolated, amplified by the polymerase chain reaction (PCR), and used in a second round of EMSA selection. Either seven (CHO preparation) or eight (baculovirus preparation) rounds of selection in total were performed before subcloning individual members of the selected sequences. As each round was expected to enrich for better binding sites, the final subcloned oligonucleotides were expected to contain high affinity binding sites for the C1, C2, and C2′ complexes. In addition, such a procedure should give some indication of the relative stringency of selection for a given base at a particular position within the binding site consensus.

The selected sequences were placed in three separate groups for analysis (Fig. 7). Group I contains sequences that were selected by the C1

complex from either mammalian or baculovirus preparations. These sequences were pooled for analysis because with both preparations formation of the C1 complex requires only Myc protein, and because the two sets of sequences (that isolated with mammalian Myc and that isolated with baculovirus Myc) were similar to each other. Most of the selected sequences in this group contained the sequence CACGTG (21/27 of sequenced subclones). By aligning all of the sequences that contained this central core sequence, it was found that the sequences flanking this core were also nonrandom. A 12 base consensus sequence of GACCACGTGCTC [SEQ ID. No. 1] was determined for sites selected by the C1 complex (see table in Fig. 7 for frequencies at each position; for a base to be included in the consensus it had to be found in at least 10 out of the 21 sequences with a CACGTG core). The C2 complex from baculovirus preparations selected sequences similar to those selected by the C1 complex (Fig. 7, Group II). Most of these selected sequences also contained the CACGTG core (19/22). These sequences had similar flanking sequences adjacent to the core hexamer to those found with the C1 complex, although there was a slight preference for GCC over CTC in the 3′ flank (see table for Group II in Fig. 7).

As expected, complexes running at the position of C2 that were selected by the mammalian Myc preparations had a greater diversity of sequences (Fig. 7, Group III). Several sequences (8/36) contained the CACGTG core. These sequences were presumably selected by the mammalian C2 complex (comprised of Myc and the 26-29 kd factor) and demonstrated the same flank preferences as the C1 complex. Several other selected sites (9/36) contained a CAGCTG core sequence presumably selected by the C2′ complex. In addition, 8 of the 36 sequences were very AT rich, and many of the sequences in all three groups contained AT rich stretches. This enrichment for AT rich sequences might reflect a preference of Myc for these sequences, or instead might simply indicate a bias arising from the protocol used. It is interesting to note, however, that

in previous filter binding experiments, the mammalian Myc preparation has demonstrated a preference for binding AT rich sequences within various plasmids or lambda genomic DNA.

To confirm the validity of our site selection procedure a number of the selected sites individually by EMSA (Fig. 8) were tested. As expected, it was found that sequences containing the core CACGTG formed both the C1 and C2 complexes (Fig. 8, probe groups 1, 2, 5, and 6) while sequences containing the CAGCTG core formed only the C2′ complex (Fig. 8, probe groups 7 and 8). Note that the C2′ activity is only present in the CHO derived Myc preparations. No complex formed when selected sequences that did not contain a canonical E box core were tested (Fig. 8 probe groups 3, 4, 9, 10, and 11). These latter sequences, therefore, do not represent high affinity sites for proteins in the Myc preparations.

Example 6

Off Rate Of The C1 And C2 Complexes

Off-rates for the Myc containing complexes were measured as a means of comparing their affinities. The off-rate of the C2 complex formed on the USE probe was approximately 1-2 minutes (Fig. 9, baculovirus Myc; similar results were obtained with CHO Myc, data not shown). The C1 complex was not fully competed in this experiment using 250 fold excess of USE competitor. Although competition was not complete, the amount of C1 complex remaining at the earliest measurable timepoint ("0") was significantly less than the starting amount and virtually equal to the maximum competition achieved in these experiments. These data are indicative of an abundant weakly binding protein with an immeasurably fast off-rate. Therefore Myc alone appears to bind significantly more weakly than does Myc and the 26-29 kd factor.

Example 7

Identification of an Inhibitor of c-Myc C2 Complex Activity in Yeast Cells

Yeast host cells are transformed with plasmids carrying a c-Myc expression vector (host 'a'); or the c-Myc expression vector and a 26-29 kilodalton C2 complex protein identified as above (host 'b'). In addition all yeast strains are cotransformed with a plasmid that contains the coding sequence for β-galactosidase operably-linked to the CACGTG sequence motif as described above.

A lawn of each of the transformed yeast strains is spread on agar plates containing X-gal in the medium and small filter disks containing compound W, X, Y, or Z are placed on the lawns. The yeast are allowed to grow and the plates are monitored for colony growth and colony color by visual observation. Typical results from such an experiment are shown in Table 1.

Table 1: Identification of Inhibitors of C2 Complex Activity

Compound	Yeast	Colony Growth	Color from β-gal Assay with X-gal
none	a	+	White
	b	+	Blue
W	a	+	White
	b	+	White
X	a	-	-
	b	-	-
Y	a	+	White
	b	+	Blue
Z	a	+	Blue
	b	+	Blue

The results of the above table indicate that compound W prevents the induction of β-galactosidease in the 'b' host cells. Therefore, compound W is an inhibitor of C2 complex hetero-oligomer formation and an inhibitor of c-Myc biological activity. Compound X inhibits the growth of yeast *per se* and thus would not be a compound of interest.

Compound Y does not prevent induction of β-galactosidase activity in the 'b' host cells. Therefore, compound Y is not an inhibitor of C2 complex hetero-oligomer formation.

Compound Z shows an interesting effect of inducing β-galactosidase activity in the 'a' host cells which does not contain the C2 complex protein used in the 'b' hosts, rather than preventing hetero-oligomer formation. This suggests that compound Z may induce synthesis of a partner protein

which is not otherwise present in the yeast host cells or that it may be (or mimic) such a protein.

From these results, compound W would be identified as an inhibitor of C2 complex formation and/or DNA binding and thus of c-Myc transcriptional activity *in vivo*.

Example 8

Identification of an Inhibitor of c-Myc C2´ Complex Activity in Yeast Cells

Yeast host cells are transformed with two plasmids, each plasmid carrying a C2´ complex expression vector encoding at least one 40-50 kilodalton C2´ peptide (host 'a'); or the c-Myc expression vector in addition to the vectors encoding the C2´ complex proteins identified as above (host 'b'). In addition all yeast strains are cotransformed with a plasmid that contains the coding sequence for β-galactosidase operably-linked to the CAGCTG sequence motif as described above.

A lawn of each of the transformed yeast strains is spread on agar plates containing X-gal in the medium and small filter disks containing compound W, X, Y, or Z are placed on the lawns. The yeast are allowed to grow and the plates are monitored for colony growth and colony color by visual observation. Typical results from such an experiment are shown in Table 1.

Table 2: Identification of Inhibitors of C2´ Complex Activity

Compound	Yeast	Colony Growth	Color from β-gal Assay with X-gal
none	a	+	White
	b	+	Blue
W	a	+	White
	b	+	White
X	a	-	-
	b	-	-
Y	a	+	White
	b	+	Blue
Z	a	+	Blue
	b	+	Blue

The results of the above table indicate that compound W prevents the induction of β-galactosidase in the 'b' host cells. Therefore, compound W is an inhibitor of C2´ complex hetero-oligomer formation and an inhibitor of the c-Myc biological activity that is directed towards promoting such C2´ complex hetero-oligomer formation. Compound X inhibits the growth of yeast *per se* and thus would not be a compound of interest.

Compound Y does not prevent induction of β-galactosidase activity in the 'b' host cells. Therefore, compound Y is not an inhibitor of C2 complex hetero-oligomer formation.

Compound Z shows an interesting effect of inducing β-galactosidase activity in the 'a' host cells which does not contain the Myc protein used

in the 'b' hosts, rather than preventing hetero-oligomer formation. This suggests that compound Z may induce synthesis of a protein that can substitute for Myc in promoting formation of the C2′ complex which is not otherwise present in the yeast host cells or that it may be (or mimic) such a protein.

From these results, compound W would be identified as an inhibitor of C2′ complex formation and/or DNa binding activity and thus of c-Myc transcriptional activity *in vivo*.

All references cited herein are fully incorporated by reference. Having now fully described the invention, it will be understood by those with skill in the art that the scope may be performed within a wide and equivalent range of conditions, parameters and the like, without affecting the spirit or scope of the invention or any embodiment thereof.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Kingston, Robert E
Papoulas, Ophelia

(ii) TITLE OF INVENTION: C-MYC DNA BINDING PARTNERS, MOTIFS, SCREENING ASSAYS, AND USES THEREOF

(iii) NUMBER OF SEQUENCES: 101

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox
(B) STREET: 1225 Connecticut Avenue, N.W., Suite 300
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20036

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Cimbala, Michele A
(B) REGISTRATION NUMBER: 33,851
(C) REFERENCE/DOCKET NUMBER: 0609.3440004

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (202)833-7533
(B) TELEFAX: (202)833-8716

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GACCACGTGC TC 12

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GACCACGTGG TC 12

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AACAGTYCTG TT 12

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCGATATCTC GAGACGG 17

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGACGATCTA GATTCG 16

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATCTCTGCA GCAGCTGGCA GCAGCTGGCA GCAGCTGGCG 40

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCTGCAGT CATGTGGCGT CATGTGGCGT CATGTGGCAG 40

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATCTGCAGT CACGTGGCGT CACGTGGCGT CACGTGGCAG 40

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TCGACGTCGC AGCAGGTGCA G 21

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCGACCCCAC CAGCTGGCGA G 21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGCTTCGAAC ACCTGCAGCA GCTGGCAGGA AGCAGGCCTA 40

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGCTTTAAAA TCCCCACCAG CTGGCGAAGC AACAGGTGCA 40

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AATTGCGAAA CCCCTGGAAT ATTCCGACCT GGCAGCCTC 39

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCGACTTTAG ACCACGTGGT CCCCTCGA 28

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGACGATCTA GATTCGNNNN NNNNNNNNNN NNNNNNCCGT CTCGAGTATC GG 52

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

NNCANNTGNN 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCAGAATCTA CCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGGGCTACCA CGTGCTTATG 20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGACGAAAGC ACGTGCTCCG 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCACATGACC ACGTGCTCTG 20

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGCAGAGACA CGTGCCCTGG 20

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGCAAACCAC GTGTTATGTG 20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGACCACGTG CTCTTCGACT TG 22

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCACAATTTG TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGACAACATC GACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCCTGCATGA CCACGTGGAC C 21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCAAATATGA CCACGTGGTA C 21

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGACCACGTG CTCTTTTGTG 20

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATAAACT CCACGTGGTC C 21

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CGGGCACGTG CTCCTCGGAC TG 22

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GGTAGCAAAA AGCACGTGCC CG 22

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GGGGGATTTA AGCACGTGCT CC 22

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CACCTATTAA CCACGTGGTA C 21

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GACCACGCGG CATCCACGTG CCGT 24

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGGGACCACG TGCTCGGTTG 20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CACATATTAG ACCACGTGCT CC 22

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CGGCCACGTG CTCACTGTCT ACC 23

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGATGGACAG CTTCTTCCTG 20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCAATCCCCC GCTGCTCGCC 20

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCCAAAAATG TACAGCTGTG CC 22

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CGGCCACGAG GTCATGAATG TGC 23

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCAGGCTGTA CGTGACTTGG 20

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CCGCAGTCCT GGTGCTCTGC 20

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CACTAAGAAA TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GGGGATTTAA GCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGGCCACGTG CCTTCTTTCT CCG 23

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CATAGTCGAG AGAGCACGTG CGAA 24

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CATAAGTCAG ACCACGTGGC CG 22

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CCCAACTAAG ACCACGTGGC CG 22

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CAGTCGAAGA GGCCACGTGG CGA 23

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGTAGGTTAT TCCCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CATAAATAGG CCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGAAAATGTA CCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GGAACAGACC ACGTGGCTTG 20

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GTACCACGTG CTTTTTTGGC 20

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CAGTCCGAGG AGCACGTGCC CG 22

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CGGCCACGTG TCGAGCATGA GTC 23

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CGGCCACGTG CTCGTAAATT TGC 23

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GCGACAAAAT TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CGCAAAATCG ACCACGTGGT CC 22

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCATAAGTAA TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GCAAAAAAAC CACGTGGTCC 20

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GGGGGCGGAA CTCCGTTGTC 20

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GGGGACCCGA TCTCTCGCTG 20

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CAATAATATT TCCTTTCCTG 20

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTCCACGCGG CATCCACGTG CCGT 24

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CGGCCACGTG CTCTATACAT GCC 23

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GGACCACGTG CTTATCTTTG 20

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CGACCACGTG TTCCGCTACT CG 22

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CGAGTAGCGA GCACGTGTTG C 21

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GCACCACGTG CTTACCATGT C 21

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GGACAAAAAG CACGTGCTAC 20

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GCAAAACTCC ACGTGGTCGG 20

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GGGCAAAAAC AACAGCTGTG CG 22

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GGGAAAGAGA TCAGCTGTGC G 21

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GGAGAATTGA ACAGCTGACC C 21

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GGGACAAACC AGTCAGCTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GGGCACAGCT GTTTAGTGGG 20

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

GGCAAGCGGA CAGCTGTTCC 20

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GGCATTGATC AGCTGTGTGG 20

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

GCAAAAACCA GCTGGTCCCC 20

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

CGCAAGTGTA ACAGCTGGTG C 21

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GGATGGTTTT TTTTTTGTAC 20

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GCATGATTTT CTTTTTGTCC 20

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

CAGAGTTTTT TTGAGCCCCC 20

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

GCAAAAAATA AAAATACATC 20

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GGCAAAAAAG TCAAAATACG 20

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GCACAATAAA AAACTTTGCG 20

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

CCATATGTTC ATTGTTGTCC 20

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

CACAAAAATT TAGTGTGTGC 20

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

CGGCCCCGTG CTCTAGCCCA TGC 23

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

CGGGGAAGTC CCAAGTGCCC C 21

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

CACAGGAACA TACACGGGCC CG 22

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GGGACGGGAT GATTGACGTG CCGT 24

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

CGCAAGCGAC GTCAGTCCTG 20

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CACCTACCAC TGATCGCGGC 20

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

GGACAAACAT CCCATTACCC 20

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

GGGGATGGAA CATCGCGCTG 20

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

CCAGTCGGGC CTAACCGGCC 20

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GGGAGCCATC GACGCCGGTG 20

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

CCATAGGGGA GTTGACAGCC 20

WHAT IS CLAIMED IS:

1. A method for the purification of Myc from a mammalian source, wherein said method comprises:

(a) growing mammalian cells capable of expressing c-Myc;

(b) inducing c-Myc expression in said cells;

(c) lysing the membrane of said mammalian cells and purifying nuclei therefrom;

(d) treating said nuclei in a buffer comprising DNase I;

(e) solubilizing said nuclei in a buffer comprising sodium dodecyl sulfate or urea at concentrations greater than 4 M and separating the nuclear pellet from the supernatant fraction;

(f) applying said supernatant fraction of step (e) to a DEAE Sepharose CL-6B column and eluting bound c-Myc from said DEAE Sepharose CL-6B column with a salt gradient;

(g) applying said c-Myc of step (f) to a FPLC Mono-Q column and eluting bound c-Myc with a salt gradient.

2. A method for the detection of C1 complexes in a sample, wherein said method comprises detecting DNA binding of c-Myc-containing homo-oligomers to the DNA motif 5'-CACGTG-3', in its double stranded DNA form.

3. A method for the detection of C2 complexes in a sample, wherein said method comprises detecting DNA binding of c-Myc-containing hetero-oligomers to the DNA motif 5'-CACGTG-3', in its double stranded DNA form.

4. A method for the detection of C2´ complexes in a sample, wherein said method comprises detecting c-Myc directed DNA binding to the DNA motif 5'-CAGCTG-3', in its double stranded DNA form.

5. A protein composition comprising at least one peptide capable of forming a C2 complex, wherein said peptide capable of forming a C2 complex is found in a 26-29 kD protein fraction purified from Chinese hamster ovary cells or baculovirus.

6. The protein composition of claim 5, wherein said protein composition is prepared from Chinese hamster ovary cells by a method comprising the steps of:

(a) growing said cells;

(b) lysing the membrane of said cells and purifying nuclei therefrom;

(c) treating said nuclei in a buffer comprising DNase I;

(d) solubilizing said nuclei in a buffer comprising sodium dodecyl sulfate or urea at concentrations greater than 4 M and separating the nuclear pellet from the supernatant fraction;

(e) applying said supernatant fraction of step (e) to a DEAE Sepharose CL-6B column and the bound C2 complex protein from said DEAE Sepharose CL-6B column with a salt gradient; and

(g) applying the eluted C2 complex protein of step (f) to a FPLC Mono-Q column and eluting bound C2 complex protein with a salt gradient.

7. The protein composition of claim 5, wherein said protein composition is prepared from baculovirus.

8. A protein composition comprising at least one peptide capable of forming a C2′ complex in the presence of c-Myc, wherein said peptide capable of forming a C2′ complex in the presence of c-Myc is found in a 40-50 kD protein fraction purified from CHO cells.

9. The protein composition of claim 8, wherein said protein composition is prepared from Chinese hamster ovary cells by a method comprising the steps of:

(a) growing said cells;

(b) lysing the membrane of said cells and purifying nuclei therefrom;

(c) treating said nuclei in a buffer comprising DNase I;

(d) solubilizing said nuclei in a buffer comprising sodium dodecyl sulfate or urea at concentrations greater than 4 M and separating the nuclear pellet from the supernatant fraction;

(e) applying said supernatant fraction of step (e) to a DEAE Sepharose CL-6B column and the bound C2′ complex protein from said DEAE Sepharose CL-6B column with a salt gradient; and

(g) applying the eluted C2′ complex protein of step (f) to a FPLC Mono-Q column and eluting bound C2′ complex protein with a salt gradient.

10. A method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity, wherein said method comprises detecting the ability of said compound to inhibit C1 complex formation, C2 complex formation or C2′ complex formation.

11. The method of claim 10, wherein said complex formation is C1 complex formation.

12. The method of claim 10, wherein said complex formation is C2 complex formation.

13. The method of claim 10, wherein said complex formation is C2´ complex formation.

14. A method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity, wherein said method comprises detecting the ability of said compound to inhibit C1 complex DNA binding, C2 complex DNA binding, or C2´ complex DNA binding.

15. The method of claim 14, wherein said DNA binding is C1 complex DNA binding.

16. The method of claim 15, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CACGTG-3'.

17. The method of claim 15, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CATGTG-3'.

18. The method of claim 14, wherein said DNA binding is C2 complex DNA binding.

19. The method of claim 18, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CACGTG-3'.

20. The method of claim 18, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CATGTG-3'.

21. The method of claim 14, wherein said DNA binding is C2´ complex DNA binding.

22. The method of claim 21, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CAGCTG-3'.

23. A method for the purification of a peptide capable of forming a C2 or C2´ complex, or a mixture of such peptides from a crude preparation, wherein said method comprises extraction of Chinese hamster ovary cells and assay of said peptide by detection of the ability of said peptide to form said C2 or said C2´ complex.

24. A method for identifying and classifying a compound as an inhibitor of c-myc hetero-oligomer DNA binding wherein said method comprises evaluating the ability of said compound to alter expression of a reporter gene in a host cell, wherein expression of said reporter gene is operably-linked to DNA binding by said hetero-oligomer to an oligonucleotide comprising the sequence 5'-CACGTG-3'.

25. The method of claim 24, wherein expression of said reporter gene induces a phenotypic change in a host cell.

26. The method of claim 24, wherein said reporter gene is *lacZ*.

27. The method of claim 24, wherein said reporter gene is CAT.

28. The method of claim 24, wherein said reporter gene is *LEU2*.

29. The method of claim 24, wherein said phenotypic change is detected by visual inspection of the host cell.

30. The method of claim 24, wherein said host is *S. cerevisiae*.

31. The method of claim 24, wherein said host is a mammalian cell.

32. A method for identifying and classifying a compound as an inhibitor of c-Myc-directed C2´ hetero-oligomer DNA binding wherein said method comprises evaluating the ability of said compound to alter expression of a reporter gene in a host cell, wherein expression of said reporter gene is operably-linked to DNA binding by said hetero-oligomer to an oligonucleotide comprising the sequence 5'-CAGCTG-3'.

33. The method of claim 32, wherein expression of said reporter gene induces a phenotypic change in a host cell.

34. The method of claim 32, wherein said reporter gene is *lacZ*.

35. The method of claim 32, wherein said reporter gene is CAT.

36. The method of claim 32, wherein said reporter gene is *LEU2*.

37. The method of claim 32, wherein said phenotypic change is detected by visual inspection of the host cell.

38. The method of claim 32, wherein said host is *S. cerevisiae*.

39. The method of claim 32, wherein said host is a mammalian cell.

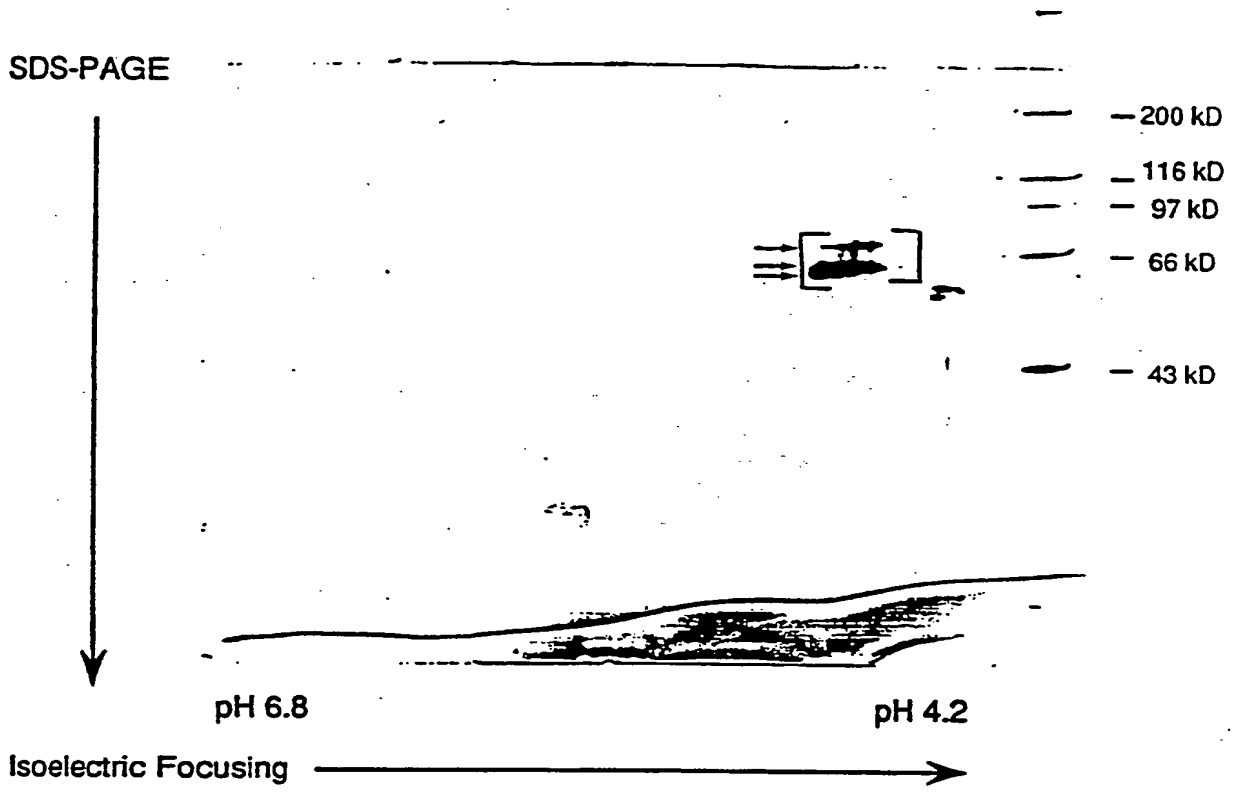

SDS-PAGE
200 kD
116 kD
97 kD
66 kD
43 kD
pH 6.8
pH 4.2
Isoelectric Focusing

FIGURE 1A

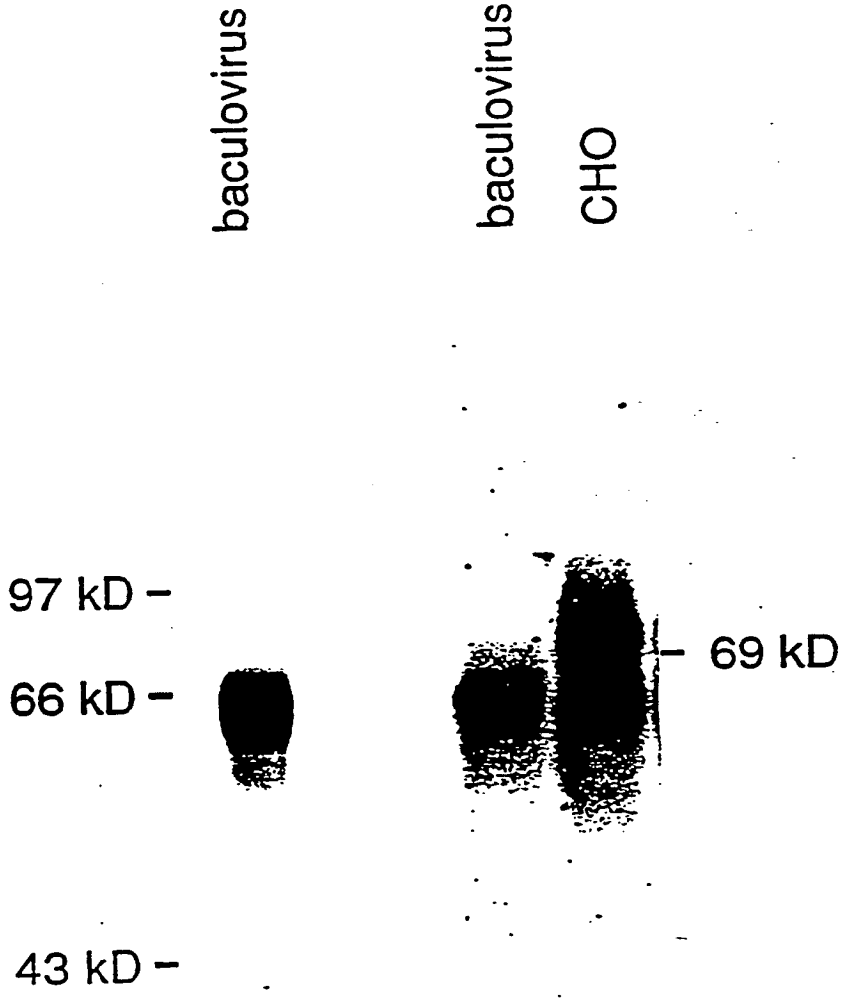
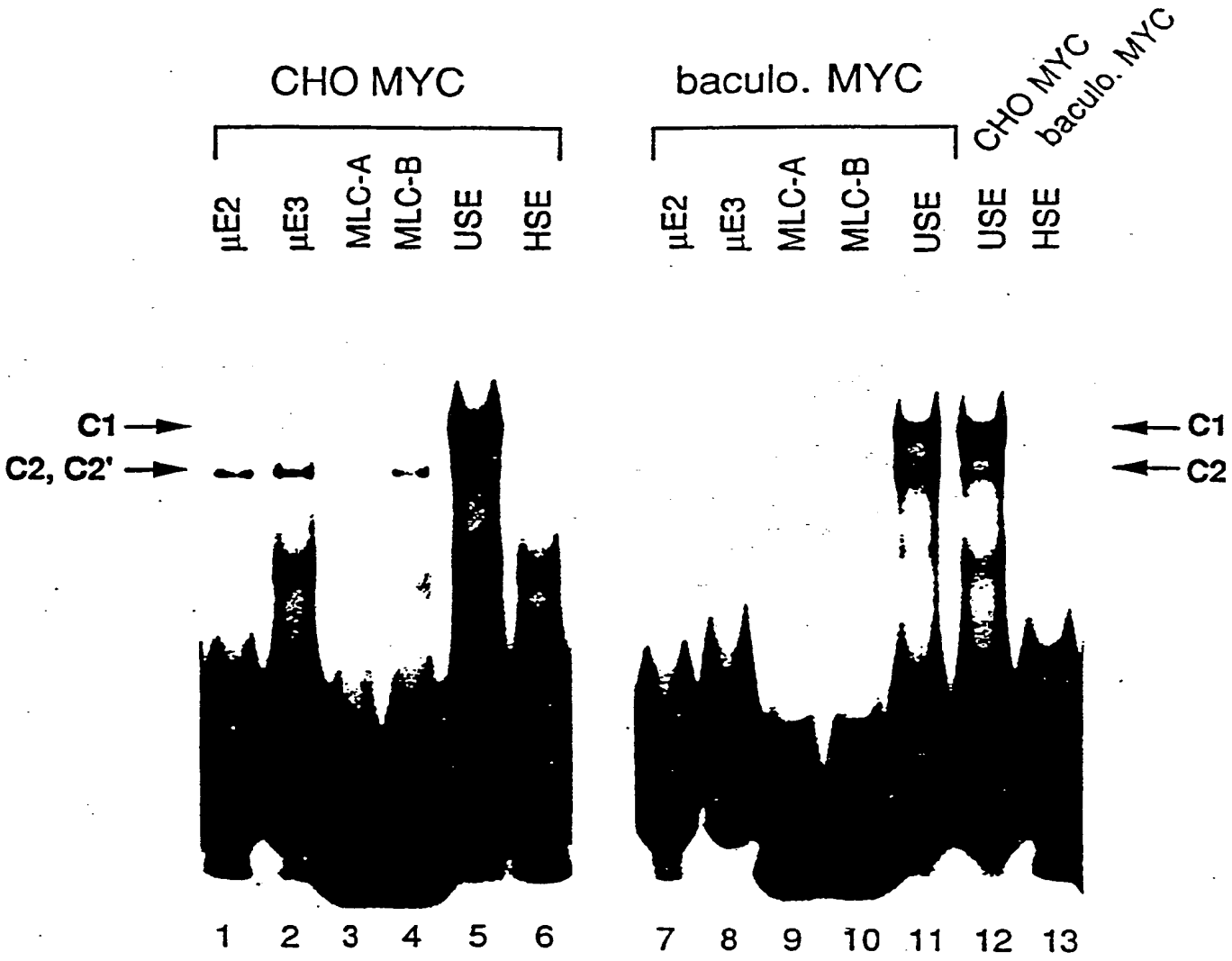

Silver Stain
Immunoblot
baculovirus
baculovirus
CHO
97 kD –
66 kD –
43 kD –
69 kD

FIGURE 1F



E-Box Probes

MLC-B (µE2)	CAGCTG
µE3	CATGTG
USE	CACGTG
MLC-A (kE2)	CAGGTG

FIGURE 2

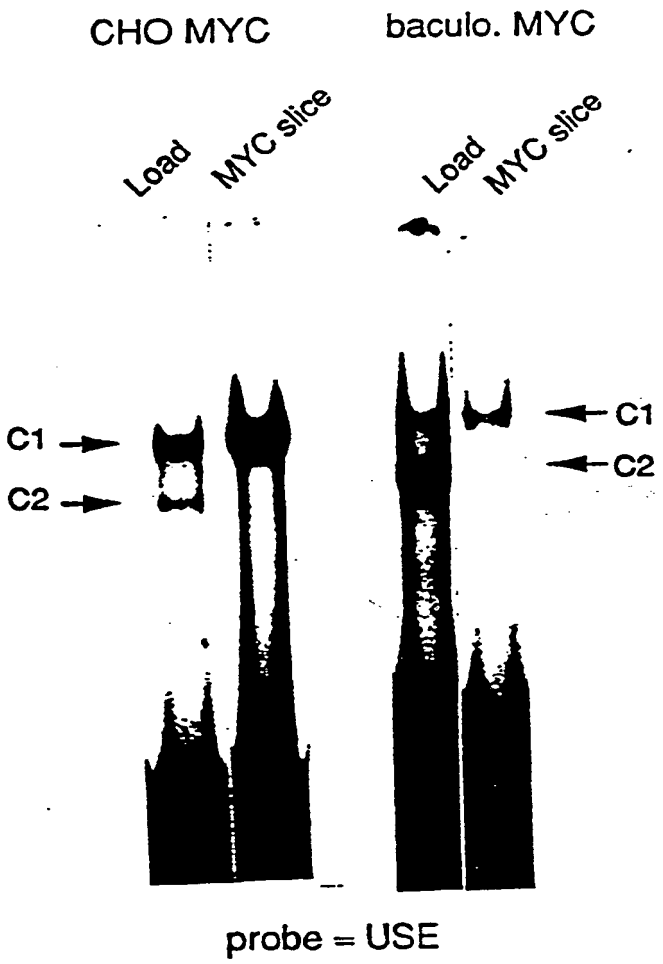

CHO MYC
baculo. MYC
Load
MYC slice
Load
MYC slice
C1
C2
C1
C2
probe = USE

FIGURE 3

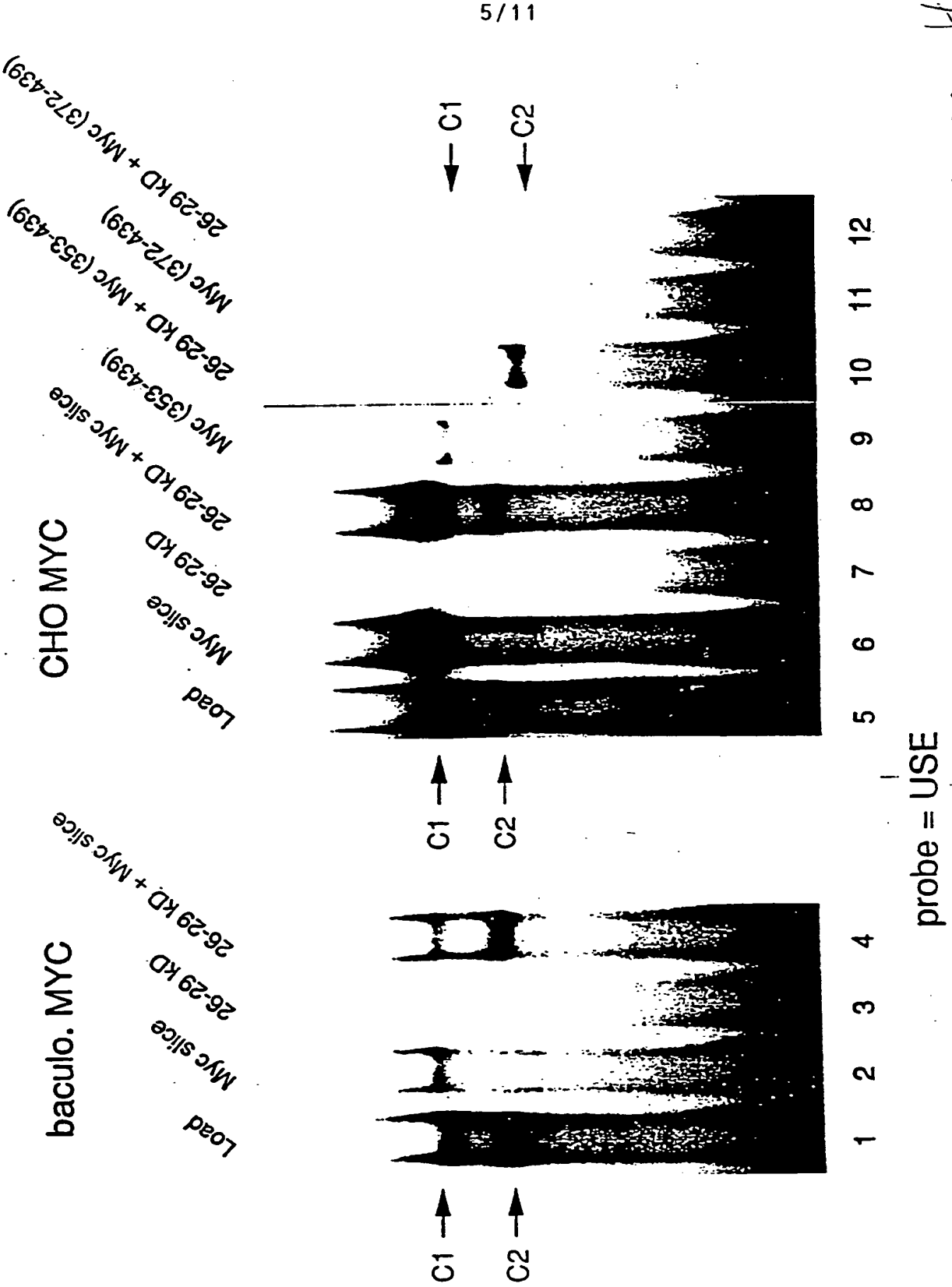

baculo. MYC
CHO MYC
load
Myc silica
26-29 kD
26-29 kD + Myc silica
Myc (353-439)
26-29 kD + Myc (353-439)
Myc (372-439)
26-29 kD + Myc (372-439)
C1
C2
1 2 3 4 5 6 7 8 9 10 11 12
probe = USE

FIGURE 4

FIGURE 5

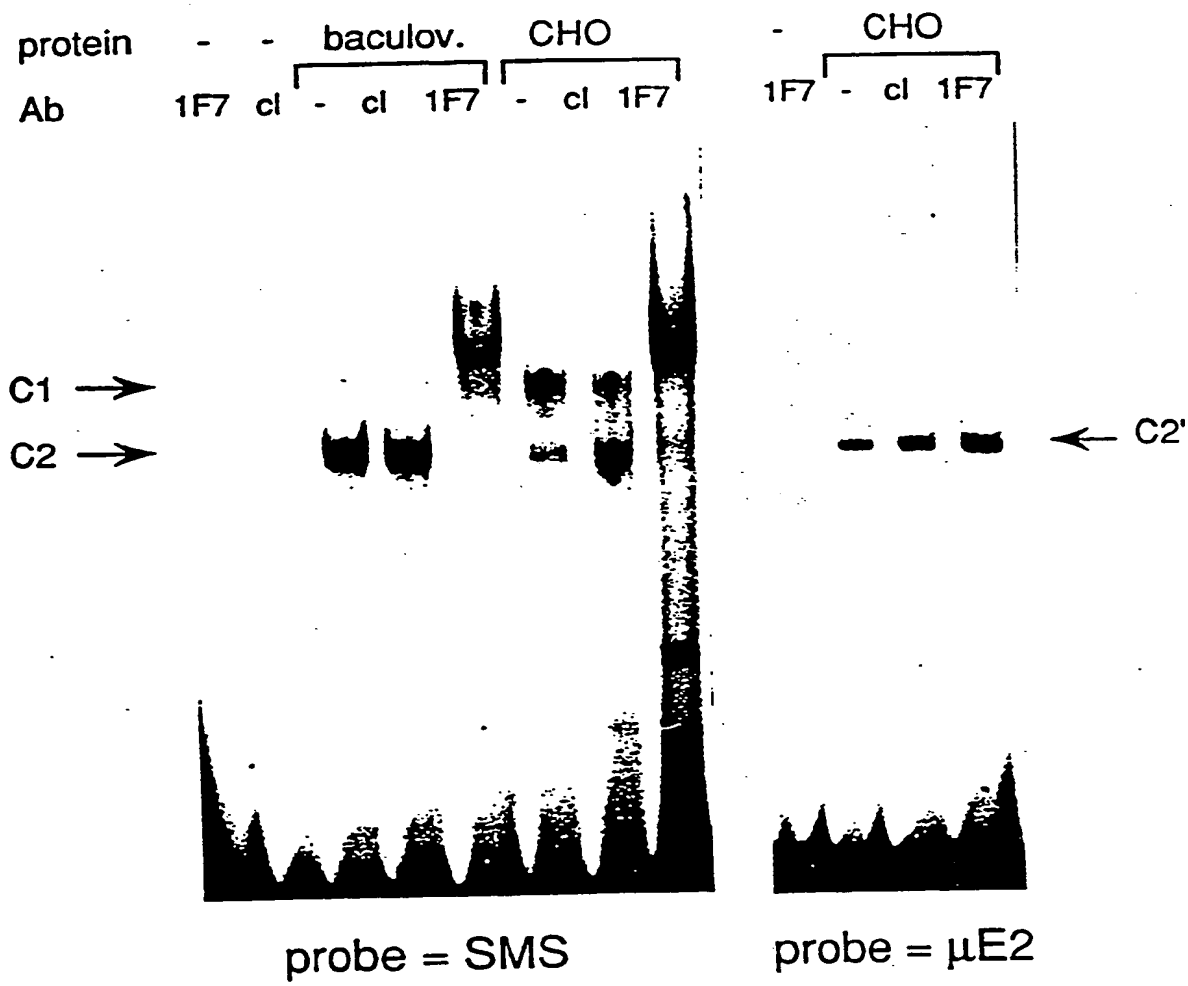

protein - - baculov. CHO - CHO
Ab 1F7 cl - cl 1F7 - cl 1F7 1F7 - cl 1F7
C1
C2
C2'
probe = SMS
probe = µE2

FIGURE 6

GROUP I: CHO AND BACULOVIRUS C1 COMPLEXES SELECT THE CACGTG CORE

SEQ ID No.					
17	GCAGAATCTAC	CACGTG	CTCC		
18	GGGGCTAC	CACGTG	CTTATG		
19	GGACGAAAG	CACGTG	CTCCG		
20	GCACATGAC	CACGTG	CTCTG	1	
21	GGCAGAGA	CACGTG	CCCTGG		
22	GGCAAAC	CACGTG	TTATGTG	*	
23	CGAC	CACGTG	CTCTTCGACTTG		
24	GCACAATTTGTAC	CACGTG	GCCG		
25	GGACAACATCGAC	CACGTG	GCCG		
26	GCCTGCATGAC	CACGTG	GACC		
27	GCAAATATGAC	CACGTG	GTAC		
28	GGAC	CACGTG	CTCTTTTGTG		
29	GGCATAAACTC	CACGTG	GTCC		
30	CGGG	CACGTG	CTCCTCGGACTG		
31	GGTAGCAAAAAG	CACGTG	CCCG	2	
32	GGGGGATTTAAG	CACGTG	CTCC		
33	CACCTATTAAC	CACGTG	GTAC		
34	GACCACGCGGCATC	CACGTG	CCGT		
35	GGGGAC	CACGTG	CTCGGTTG		
36	CACATATTAGAC	CACGTG	CTCC		
37	CGGC	CACGTG	CTCACTGTCTACC		
38	GGATGGA	CAGCTT	CTTCCTG		
39	GCAATCCC	CCGCTG	CTCGCC	3	
40	GCCAAAAATGTA	CAGCTG	TGCC		
41	CGGC	CACGAG	GTCATGAATGTGC		
42	GCAGGCTG	TACGTG	ACTTGG		
43	CCGCAGTC	CTGGTG	CTCTGC		

	-3	-2	-1	CACGTG	1	2	3
C	1	0	16		14	5	16
G	10	3	4		6	0	1
A	7	16	1		0	1	3
T	3	2	0		1	15	1

GROUP II: BACULOVIRUS C2 COMPLEXES SELECT THE CACGTG CORE

Seq ID No.					
44	CACTAAGAAATAC	CACGTG	GCCG		
45	GGGGATTTAAG	CACGTG	CTCC		
46	CGGC	CACGTG	CCTTCTTTCTCCG		
47	CATAGTCGAGAGAG	CACGTG	CGAA		
48	CATAAGTCAGAC	CACGTG	GCCG		
49	CCCAACTAAGAC	CACGTG	GCCG	*	
50	CAGTCGAAGAGGC	CACGTG	GCGA		
51	CGTAGGTTATTCC	CACGTG	GCCG		
52	CATAAATAGGC	CACGTG	CTCC	*	
53	GGAAAATGTAC	CACGTG	CTCC		
54	GGAACAGAC	CACGTG	GCTTG		
55	GTAC	CACGTG	CTTTTTTGGC		
56	CAGTCCGAGGAG	CACGTG	CCCG		
57	CGGC	CACGTG	TCGAGCATGAGTC		
58	CGGC	CACGTG	CTCGTAAATTTGC		
59	GCGACAAAATTAC	CACGTG	GCCG		
60	CGCAAAATCGAC	CACGTG	GTCC		
61	GCATAAGTAATAC	CACGTG	GCCG		
62	GCAAAAAAAC	CACGTG	GTCC		
63	GGGGGCGGAACTCCGTTGTC				4
64	GGGGACCCGATCTCTCGCTG				
65	CAATAATATTTCCTTTCCTG				

	-3	-2	-1	CACGTG	1	2	3
C	0	1	16		8	11	13
G	11	5	3		10	1	2
A	2	13	0		0	0	1
T	6	0	0		1	7	3

FIGURE 7A

GROUP III: THREE CLASSES OF SITES ARE SELECTED BY CHO COMPLEXES MIGRATING AS C2 AND C2'

Seq ID No

USE CORE (CACGTG)

66	GTCCACGCGGCATC	CACGTG	CCGT
67	CGGC	CACGTG	CTCTATACATGCC
68	GGAC	CACGTG	CTTATCTTTG 5
69	CGAC	CACGTG	TTCCGCTACTCG *
70	CGAGTAGCGAG	CACGTG	TTGC
71	GCAC	CACGTG	CTTACCATGTC
72	GGACAAAAAG	CACGTG	CTAC.
73	GCAAAACTC	CACGTG	GTCGG 6

	-3	-2	-1	CACGTG	1	2	3
C	2	0	6		5	1	3
G	4	1	2		1	0	2
A	2	5	0		0	0	1
T	0	2	0		2	7	2

µE2 CORE (CAGCTG)

Seq ID No.

GGGCAAAAACAA	CAGCTG	TGCG	74
GGGAAAGAGAT	CAGCTG	TGCG	75
GGAGAATTGAA	CAGCTG	ACCC	76
GGGACAAACCAGT	CAGCTG	GCCG	77
GGGCA	CAGCTG	TTTAGTGGG	78
GGCAAGCGGA	CAGCTG	TTCC 7	79
GGCATTGAT	CAGCTG	TGTGG *	80
GCAAAAAC	CAGCTG	GTCCCC 8	81
CGCAAGTGTAA	CAGCTG	GTGC	82

At-rich

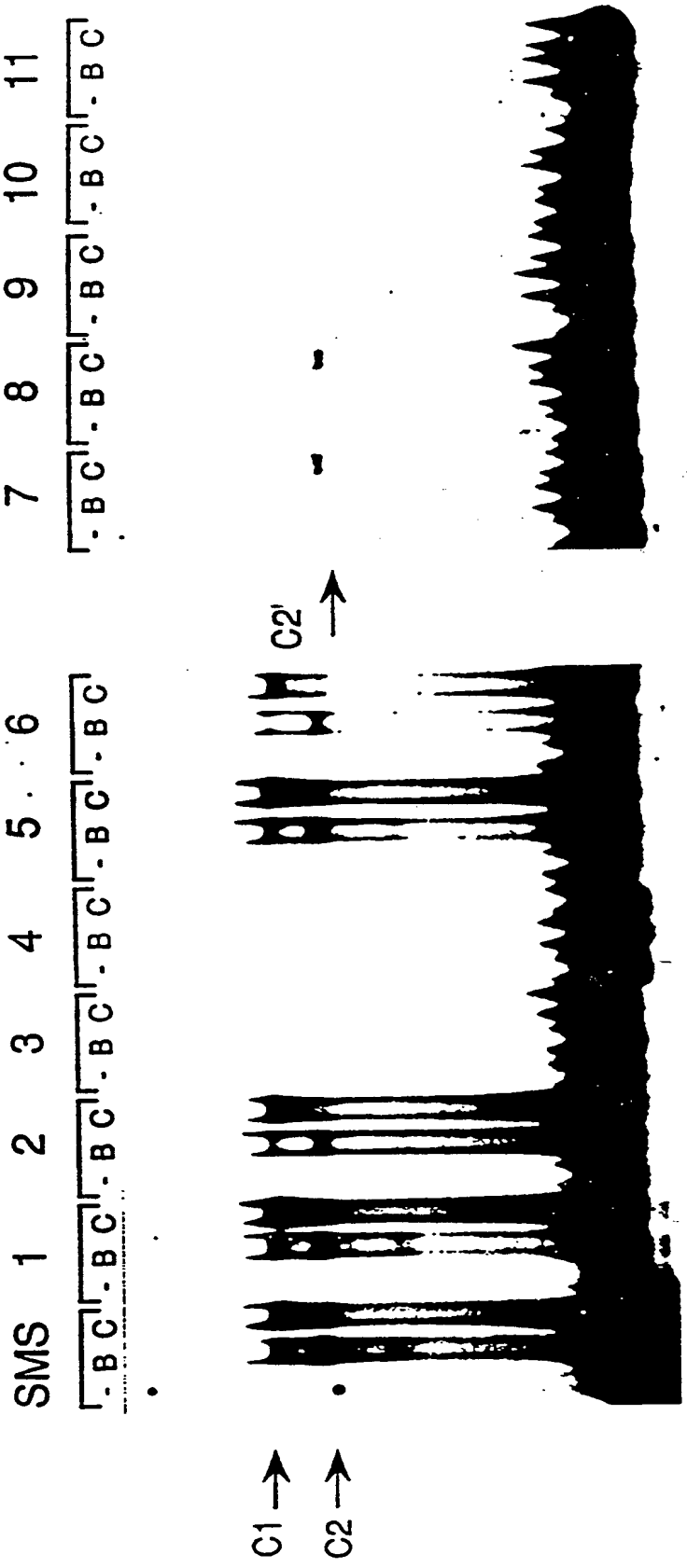
GGATGGTTTTTTTTTTGTAC		83
GCATGATTTTCTTTTTGTCC		84
CAGAGTTTTTTTGAGCCCCC		85
GCAAAAAATAAAAATACATC		86
GGCAAAAAAGTCAAAATACG		87
GCACAATAAAAAACTTTGCG	9	88
CCATATGTTCATTGTTGTCC		89
CACAAAAATTTAGTGTGTGC	10	90

Other

CGGC	CCCGTG	CTCTAGCCCATGC	91
CGGGGAAGTCC	CAAGTG	CCCC *	92
CACAGGAACATA	CACGGG	CCCG	93
GGGACGGGATGATT	GACGTG	CCGT	94
CGCAAGC	GACGTC	AGTCCTG	95
CACCTACCACTGATCGCGGC		11	96
GGACAAACATCCCATTACCC		*	97
GGGGATGGAACATCGCGCTG		*	98
CCAGTCGGGCCTAACCGGCC			99
GGGAGCCATCGACGCCGGTG			100
CCATAGGGGAGTTGACAGCC			101

FIGURE 7B

FIGURE 8



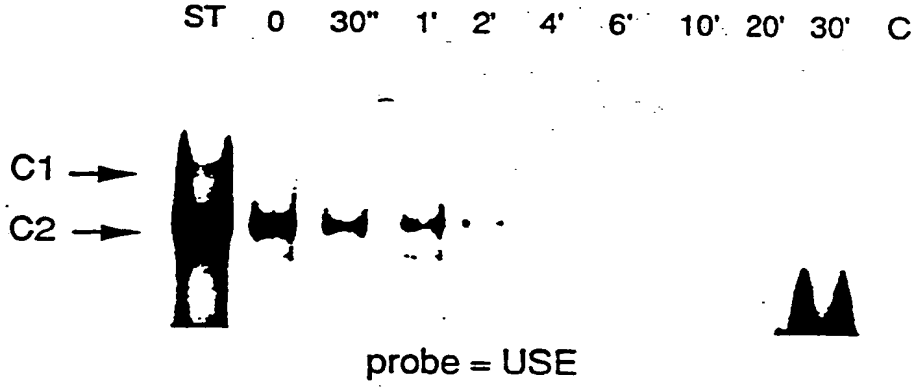


FIGURE 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :A23J 1/00; C07K 3/00; C12Q 1/68; G01N 1/00
US CL :530/412, 417; 435/6; 530/350; 424/2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/412, 417; 435/6; 530/350; 424/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Prior art disclosure of priority document.

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS
DIALOG (Biochem)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCIENCE, Volume 250, issued 23 November 1990, T.K. Blackwell, *et alii*, "Sequence-Specific DNA Binding by the c-Myc Protein", pages 1149-1151; see pages 1151, column 1.	2-4, 14-24, 32-39
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES UNITED STATES OF AMERICA, Volume 81, issued December 1984, G. Ramsey *et alii*, "The Protein Encoded by the Human Proto-Oncogene c-*myc*", pages 7742-7746, see abstract; page 7744, second column; and page 7745 first column.	10, 11
X	SCIENCE, Volume 225, issued 17 August 1984, H. Persson *et alii*, "Nuclear Localization and DNA Binding Properties of a Protein Expressed by Human c-*myc* Oncogene", pages 718-721, see page 719, first column.	8,9
X	MOLECULAR AND CELLULAR BIOLOGY, Volume 5, number 3, issued March 1985, R.A. Watt *et alii*, "Expression and Characterization of the Human c-*myc* DNA Binding Protein", pages 448-456, see page 450, first column; page 451, second column; and page 455, first column.	1-4, 23

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 09 DECEMBER 1992	Date of mailing of the international search report 11 JAN 1993
Name and mailing address of the ISA/ Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer MARGARET MOSKOWITZ Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, UNITED STATES OF AMERICA, Volume 83, issued August 1986, F. M. Wurm, "Inducible overproduction of the mouse c-*myc* protein in mammalian cells", pages 5414-5418, see page 5416, second column.	24, 25, 29, 31, 32, 37, 39
X	NATURE, Volume 296, issued 18 March 1982, P. Donner, *et alii*, "Nuclear Localization and DNA Binding of the Transforming Gene Product of Avian Myelocytomatosis Virus", pages 262-266, see page 262, second column; and page 263.	5-9
X	ANN. REV. GENET., Volume 120, issued 1986, M.D. Cole, "The *myc* Oncogene: Its Role in Transformation and Differentiation", pages 361-384; see the entire document.	10-21, 24-39

Form PCT/ISA/210 (continuation of second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1 and 23, drawn to a method of purifying Myc, classified class 530, subclass 412 and 417, and claims 5-9, drawn to a protein classified in class 530, subclass 350.

Group II, claims 2-4, drawn to a method of detection, classified in class 435, subclass 6.

Group III, claims 10-22 and 24-39, drawn to a method of identifying and/or classifying compounds, said claims are classified in class 424, subclass 2.

The inventions listed as Groups I and II do not meet the requirements for Unity of Invention for the following reasons: The invention of Group I is considered to form an inventive concept wherein said first method is directed to a purification of Myc and the product is at least one peptide capable of forming a C2 complex and that said peptide is found in a 26-29 kD protein fraction purified from Chinese hamster ovary cells or baculovirus. The invention of Group II is drawn to different methodologies, requiring different method steps and resulting in a different end product. In particular, the invention of Group II is directed to the detection of C1 complexes.

Additionally, the peptide of claim 5, "a protein composition comprising at least one peptide capable of forming a C2 complex "is admittedly and agreeably old in the art. As can be seen on page 1, bridging to page 2 of the disclosure, myc has long been known in the art as well as its sequence. Also, the prior art of record clearly demonstrates that myc has been expressed under recombinant conditions. Resultantly, the peptide of claims 5-9 cannot be considered as a special technical feature. Lacking such status, the holding of a lack of unity is justified.

The inventions listed as Groups I and III do not meet the requirements for Unity of Invention for the following reasons: As shown *supra*, the invention of Group I is comprised of a first method of making and a product made where said first method is directed to the purification of Myc and the product is at least one peptide capable of forming a C2 complex and that said peptide is found in a 26-29 kD protein fraction purified from Chinese hamster ovary cells or baculovirus. The invention of Group III is directed to "[a] method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity" (claims 10-22) and "[a] method for identifying and classifying a compound as an inhibitor of c-myc hereto-oligomer DNA binding" (claims 24-39). As can be clearly seen, the inventions of each group are drawn to different methodologies, requiring different method steps and resulting in different end products.

The inventions listed as Groups II and III do not meet the requirements for Unity of Invention for the following reasons: The inventions of Group II is drawn to a method of detection of C1 complexes while the invention of Group III is drawn to "[a] method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity" (claims 10-22) and "[a] method for identifying and classifying a compound as an inhibitor of c-myc hetero-oligomer DNA binding" (claims 24-39). Clearly said Groups are drawn to methodologies that each require different method steps and result in different end products.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims. (Telephone Practice)

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any addtional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention frist mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

CORRECTED VERSION *

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
A23J 1/00, C07K 3/00 C12Q 1/68, G01N 1/00	A1	WO 93/08701
		(43) International Publication Date: 13 May 1993 (13.05.93)

(21) International Application Number: PCT/US92/08603

(22) International Filing Date: 9 October 1992 (09.10.92)

(30) Priority data:
07/785,567 30 October 1991 (30.10.91) US

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; Fruit Street, Boston, MA 02114 (US).

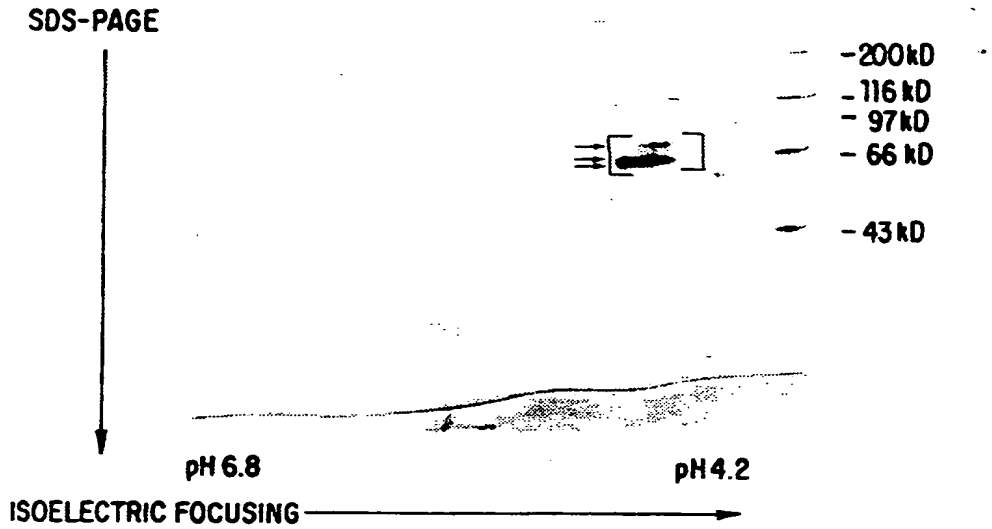
(72) Inventors: KINGSTON, Robert, E. ; 32 Draper Avenue, Arlington, MA 02174 (US). PAPOULAS, Ophelia ; 10 Emerson Place (15-B), Boston, MA 02114 (US).

(74) Agents: GOLDSTEIN, Jorge, A. et al.; Sterne, Kessler, Goldstein & Fox, 1225 Conneticut Avenue, N.W., Suite 300, Washington, DC 20036 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: C-MYC DNA BINDING PARTNERS, MOTIFS, SCREENING ASSAYS, AND USES THEREOF



(57) Abstract

The development of a mammalian cell line that overexpresses Myc and the purification of significant quantities of c-Myc from these cells is described. Three types of c-Myc-driven protein oligomerization (or complex) formations are described: (1) homo-oligomer complexes (herein termed C1 complexes) formed by association of at least two peptides of c-Myc, (2) hetero-oligomer complexes (herein termed C2 complexes) formed by heterodimerization of at least two peptides, at least one of which is not the c-Myc peptide, and specifically hetero-oligomerization between c-Myc and a 26-29 kd factor, and (3) c-Myc-dependent hetero-oligomeric complexes (herein termed the C2' complex) formed in the presence of c-Myc, however such hetero-oligomeric proteins not containing any peptides which are c-Myc. The invention is further directed to a reliable and accurate method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NL**	Netherlands
BE	Belgium	**GN**	Guinea	**NO**	Norway
BF	Burkina Faso	**GR**	Greece	**NZ**	New Zealand
BG	Bulgaria	**HU**	Hungary	**PL**	Poland
BJ	Benin	**IE**	Ireland	**PT**	Portugal
BR	Brazil	**IT**	Italy	**RO**	Romania
CA	Canada	**JP**	Japan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SK**	Slovak Republic
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
CS	Czechoslovakia	**LK**	Sri Lanka	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**MC**	Monaco	**UA**	Ukraine
DK	Denmark	**MG**	Madagascar	**US**	United States of America
ES	Spain	**ML**	Mali	**VN**	Viet Nam
FI	Finland	**MN**	Mongolia		

TITLE OF THE INVENTION

C-MYC DNA BINDING PARTNERS, MOTIFS, SCREENING ASSAYS, AND USES THEREOF

Cross-Reference to Related Applications

This application is a continuation-in-part of U.S. Application No. 07/510,253, filed April 19, 1990.

Field of the Invention

This invention is directed to methods for the purification of mammalian Myc protein, and methods for the identification of compounds that inhibit c-Myc transcriptional activity.

BACKGROUND OF THE INVENTION

Myc is a nuclear oncogene whose aberrant expression is associated with many different types of human cancers in many different tissues (Cole, M.D., *Ann. Rev. Genet. 20*:361-384 (1986)). While the mechanism of c-Myc oncoprotein action remains unknown, it clearly plays a role in the control of cell growth and differentiation (Lüscher and Eisenman, *Genes & Dev. 4*:2025-2035 (1990); Penn *et al.*, *Sem. Cancer Biol. 1*:69 (1990)). One plausible mechanism of Myc action is as a regulator of transcription in a pathway directly controlling proliferation and differentiation. This model is consistent with several observations. First, Myc has long been known as a nuclear protein with a general affinity for DNA (Abrams *et al.*, *Cell 29*:427-439 (1982); Alitalo *et al.*, *Nature 306*:274-277 (1983); Donner *et al.*, *Nature 296*:262-265 (1982); Persson and Leder, *Science 225*:718-721 (1984)), and recently a site has been identified which is specifically bound by bacterially expressed variants of c-Myc (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Prendergast and Ziff, *Science 251*:186-189 (1991)). Second, full length c-Myc has been shown to both activate and repress genes in transient transfection assays (Kaddurah-Daouk *et al.*, *Genes & Dev. 1*:347-357 (1987); Yang *et al.*, *Mol. Cell. Biol. 11*:2291-2295

(1991)), and will weakly stimulate transcription when fused to a heterologous DNA-binding domain (Lech *et al.*, *Cell 52*:179-184 (1988); Kato *et al.*, *Mol. Cell. Biol. 10*:5914-5920 (1990)). And finally, sequence similarities described below place Myc in the company of known transcription factors.

Myc contains two domains that suggest it oligomerizes, perhaps as a dimer, and binds specifically to DNA: a leucine zipper domain and a basic-helix-loop-helix (B-HLH) domain. The leucine zipper is an α-helical structure found in sequence specific DNA-binding proteins such as Fos and Jun where it mediates homo- or heterodimerization via a coiled-coiled interaction (Landschulz *et al.*, *Science 240*:1759-1764 (1988); O'Shea *et al.*, *Science 243*:538-542 (1989); and reviewed in Busch and Sassone-Corsi, *TIG 6*:36-40 (1990)). This dimerization is necessary for DNA binding (Gentz *et al.*, *Science 243*:1695-1699 (1989); Halazonetis *et al.*, *Cell 55*:917-924 (1988); Kouzarides and Ziff, *Nature 336*:646-651 (1988); Turner and Tjian, *Science 243*:1689-1694 (1989)). The HLH region also appears to mediate oligomerization necessary for DNA binding in several developmentally important proteins (Murre *et al.*, *Cell 58*:537-544 (1989); Murre *et al.*, *Cell 56*:777-783 (1989)). HLH proteins form a large and growing family and include the products of the *achaete-scute* and *daughterless* genes responsible for neural development in *Drosophila*, the R gene family which regulates pigment pattern in corn, MyoD and several other proteins involved in muscle specific differentiation in vertebrates, and a centromere binding protein, CBF1, from yeast (Braun *et al.*, *EMBO J. 8*:701-709 (1989); Cai and Davis, *Cell 61*:437-446 (1990); Caudy *et al.*, *Cell 55*:1061-1067 (1988); Cronmiller *et al.*, *Genes & Dev. 2*:1666-1676 (1988); Davis *et al.*, *Cell 51*:987-1000 (1987); Edmondson and Olson, *Genes & Dev. 3*:628-640 (1989); Ludwig and Wessler, *Cell 62*:849-851 (1990); Pinney *et al.*, *Cell 53*:781-793 (1988); Rhodes and Konieczny, *Genes & Dev. 3*:2050-2061 (1989); Villares and Cabrera, *Cell 50*:415-424 (1987); Wright *et al.*, *Cell 56*:607-617 (1989)). While many proteins

contain either an HLH or leucine zipper motif, Myc is one of a smaller number of proteins which contain both an HLH and a leucine zipper. Both the leucine zipper containing proteins and the HLH proteins require a stretch of basic amino acids adjacent to the dimerization motif to contact DNA (reviewed in Busch and Sassone-Corsi, *TIG 6*:36-40 (1990); Jones, N., *Cell 61*:9-11 (1990)). Interestingly, all B-HLH proteins appear to bind to closely related DNA sequences known as E-Boxes. These are sequence motifs found in the immunoglobulin and other tissue specific enhancers having a core of NNCANNTGNN [SEQ ID No. 16] where different central bases are preferred by different B-HLH proteins and the flanking bases can affect binding affinity (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Blackwell and Weintraub, *Science 250*:1104-1110 (1990)). The core of the reported binding site for c-Myc, CACGTG, fits this pattern and has the same core sequence as the upstream sequence element (USE) of the Adenovirus major late promoter (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Prendergast and Ziff, *Science 251*:186-189 (1991)). A cellular transcription factor (USF or MLTF) which binds to the USE has recently been cloned and also contains a B-HLH domain adjacent to a leucine zipper (Gregor *et al.*, *Genes & Dev. 4*:1730-1740 (1990)).

Many of these B-HLH or leucine zipper proteins have been found to form not only homodimers but heterodimers with other proteins having like dimerization motifs (reviewed in Busch and Sassone-Corsi, *TIG 6*:36-40 (1990); Jones, N., *Cell 61*:9-11 (1990)). Heterodimerization between specific groups of B-HLH or leucine zipper proteins can alter their DNA binding properties. While homodimers might bind weakly, heterodimers with the appropriate partner can bind with increased affinity and in some cases with a new specificity (Jones, N., *Cell 61*:9-11 (1990); Blackwell and Weintraub, *Science 250*:1104-1110 (1990); Wright *et al.*, *Mol. Cell. Biol. 11*:4104-4110 (1991)). Myc is capable of forming a homo-oligomer at high concentrations *in vitro* (Dang *et al.*, *Nature 337*:664-666 (1989); Kerkhoff and Bister, *Oncogene 6*:93-102 (1991)), although it is not clear whether

that homo-oligomer actually forms *in vivo* (Dang *et al.*, *Mol. Cell. Biol. 11*:954-962 (1991)). It seems likely that Myc directly interacts with other cellular protein(s) to form hetero-oligomer(s), and indeed one such "partner" protein, designated Max, has recently been identified (Blackwood and Eisenmann, *Science 251*:1211-1217 (1991)). The effect that such partner proteins have on Myc DNA-binding specificity is likely to be central to understanding the function of Myc.

Much of the *in vitro* work done on B-HLH proteins has utilized *in vitro* transcribed and translated protein or has used protein overexpressed in bacteria. Myc expressed by these means has been used to determine binding specificity and to demonstrate that Myc can form heterodimers with Max (Blackwell *et al.*, *Science 250*:1149-1151 (1990); Prendergast and Ziff, *Science 251*:186-189 (1991); Blackwood and Eisenmann, *Science 251*:1211-1217 (1991)). Myc, however, is post-translationally modified by at least phosphorylation in mammalian cells (Hann and Eisenmann, *Mol. Cell. Biol. 4*:2486-2497 (1984); Ramsay *et al.*, *Proc. Natl. Acad. Sci. USA 81*:7742-7746 (1984)), and post-translational modifications are believed to regulate the function of many proteins, including the transcription factors Myb, Fos, HSF, CREB, and SP-1 (Abate *et al.*, *Science 249*:1157-1161 (1990); Jackson *et al.*, *Cell 63*:155-165 (1990); Lüscher *et al.*, *Nature 344*:517-522 (1990); Sorger *et al.*, *Nature 329*:81-84 (1987); Yamamoto *et al.*, *Nature 334*:494-498 (1988)). In addition, Myc produced in avian cells has been reported to bind more tightly to DNA cellulose than bacterially produced Myc (Kerkhoff and Bister, *Oncogene 6*:93-102 (1991)).

Several lines of evidence argue that the biochemical function(s) of Myc will be determined in large part by hetero-oligomerization with Max and perhaps with other, as yet unidentified, factors. A complete understanding of the function of c-Myc will therefore require the identification of all partner proteins and a functional characterization of the complexes that these proteins form in the absence or presence of c-Myc.

To unravel the nature of Myc's function it will be necessary to determine not only the binding properties of all relevant complexes but to ascertain how they differ in action once bound. Post-translational modification might play a role in modulating the formation, binding, or further activities of these complexes and the availability of large quantities of modified c-Myc, such as described here, should facilitate a biochemical approach to this problem. Such studies should lead us to an understanding of the complexes available at different times in different cell types and the consequences for each cell in terms of appropriate growth and differentiation, or oncogenesis.

Further, to date, no inhibitors of c-Myc action have been identified. The identification of such inhibitors has suffered for lack of identification of a specific DNA binding sequence to which c-Myc binds, and for lack of a simple, inexpensive and reliable screening assay which could rapidly identify potential inhibitors and active derivatives thereof. Thus a need also still exists for rapid, economical screening assays which identify specific inhibitors of c-Myc activity.

SUMMARY OF THE INVENTION

Recognizing the potential importance of inhibitors of c-Myc oncoprotein activity in the therapeutic treatment of many forms of cancer, and cognizant of the lack of a simple assay system in which such inhibitors might be identified, the inventors have investigated c-myc DNA binding.

These efforts led to the development of a mammalian cell line that overexpresses Myc and the purification of significant quantities of c-Myc from these cells. These efforts culminated in the discovery of three types of c-Myc-driven protein oligomerization (or complex) formations: (1) homo-oligomer complexes (herein termed C1 complexes) formed by association of at least two peptides of c-Myc, (2) hetero-oligomer complexes (herein termed C2 complexes) formed by heterodimerization of

at least two peptides, at least one of which is not the c-Myc peptide, and specifically hetero-oligomerization between c-Myc and a 26-29 kd factor, and (3) c-Myc-dependent hetero-oligomeric complexes (herein termed the C2′ complex) formed in the presence of c-Myc, however such hetero-oligomeric proteins not containing any peptides which are c-Myc.

Accordingly, the invention is directed to a reliable and accurate method for the purification of Myc from a mammalian source.

The invention is further directed to the use of oligomers containing the DNA motif 5'-CACGTG-3', in its double stranded DNA form, as a reliable and accurate method for the detecfion of the presence of C1 complexes in a sample.

The invention is further directed to the use of the DNA motif 5'-CACGTG-3', in its double stranded DNA form, as a reliable and accurate method for the detection of C2 complexes in a sample.

The invention is further directed to the use of the DNA motif 5'-CAGCTG-3', in its double stranded DNA form, as a reliable and accurate method for the detection of C2′ complexes in a sample.

The invention is further directed to a 26-29 kD protein fraction purified from Chinese hamster ovary (CHO) cells or baculovirus, such protein fraction containing at least one peptide capable of forming C2 complex oligomers with c-Myc.

The invention is further directed to a 40-50 kD protein fraction purified from CHO cells, such protein fraction containing at least one peptide capable of forming C2′ complex oligomers in the presence of c-Myc.

The invention is further directed to a reliable and accurate method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity, and especially as an inhibitor of C1 complex formation, C2 complex formation or C2′ complex formation.

The invention is further directed to a reliable and accurate method for objectively classifying compounds, including human pharmaceuticals, as

inhibitors of c-Myc activity, and especially as an inhibitor of C1 complex DNA binding, C2 complex DNA binding, or C2′ complex DNA binding.

The invention further provides a method for identifying and classifying the mechanism of action of a bioactive c-Myc-inhibiting compound.

The invention further provides an assay for the monitoring of the isolation and/or purification of a peptide capable of forming a C2 or C2′ complex, or a mixture of such peptides from a crude preparation.

The invention further provides an assay for the monitoring of the isolation and/or purification of an c-Myc-inhibiting compound or mixture of such compounds from a crude preparation of such compounds.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. Purified c-Myc Protein. A) 1 μg of c-Myc protein purified from the 5A overexpressing CHO cell line was subjected to 2-dimensional gel electrophoresis. An isoelectric focusing tube gel was run with pH 5-7 ampholytes followed by SDS-PAGE and silver staining. The Myc proteins are bracketed and arrows distinguish the 60, 62, and 72 kD species. The gel was trimmed for this figure; the actual pI range for the Myc proteins was 5.0-5.6. B) 0.5 μg of purified c-Myc protein from the indicated cell lines was electrophoresed on an SDS gel and either visualized by silver staining (left lane) or electroblotted to nitrocellulose and subjected to immunoblotting using the ST-2 polyclonal antibody (right 2 lanes).

Fig. 2. DNA Binding of Purified c-Myc Proteins. The EMSA was carried out as described in materials and methods using equal amounts (approximately 2 ng) of the following probes and 0.5 μg of either purified CHO produced c-Myc or baculovirus produced c-Myc: $(\mu E2)_3$ lanes 1 and 7, $(\mu E3)_3$ lanes 2 and 8, MLC-A lanes 3 and 9, MLC-B lanes 4 and 10, $(USE)_3$ lanes 5, 11, and 12, and HSE lanes 6 and 13. Full probe

sequences are given in materials and methods. Lanes 1-6 and lanes 7-13 are different exposures of lanes from the same gel.

Fig. 3. C1 Binding Activity is Present in Myc containing Slices of SDS Gels. 400 μg of CHO produced c-Myc or 163 μg of baculovirus produced c-Myc was separated on an SDS-PAGE gel. Proteins from 0.5 cm slices were recovered, renatured as described in materials and methods, and analyzed by EMSA using the (USE)3 probe. 0.4 μg of the CHO Myc load and 5 μl of the protein from the CHO Myc-containing slice were analyzed on the same gel (left panel). 0.37 μg of the baculovirus Myc load and 5 μl of the protein from the baculovirus Myc slice were analyzed on the same gel (right panel). Slices from other molecular weight ranges of the same gel showed no binding (data not shown).

Fig. 4. Activity is Formed by c-Myc and a 26-29 kD Factor. Proteins from gel slices were recovered and analyzed by EMSA as described in materials and methods using the $(USE)_3$ probe. Lanes 1-4 represent proteins from the same gel loaded with baculovirus produced Myc described for Fig. 5. These lanes contain 0.37 μg of the loaded material (lane 1), 0.75 μg BSA with 7.5 μl of proteins from either a Myc slice (lane 2) or a 26-29 kD slice (lane 3), or 7.5 of each slice used for lanes 1 and 2 plus 0.2 μg of BSA (lane 4). Lanes 5-8 and 10 contain proteins from gels loaded with Myc purified from CHO cells. These lanes contain 0.47 of the gel load (lane 5), 4 μl of material from a Myc slice of a gel loaded with 400 μg of Myc (lane 6), 7 μl of material from a 26-29 kD slice of a similar gel plus 0.8 μg Protein A (lane 7), and both 4 μl of the Myc slice and 7 μl of the 26-29 kD slice (lane 8). Lanes 9-12 utilize the bacterially expressed Protein A-Myc fusion proteins containing either the Myc B-HLH and leucine zipper domains (amino acids 353-439) or lacking the basic region and containing Myc amino acids 372-439. These were expressed and purified as described in materials and methods. Lane 9 contains 0.5 μg of Protein A-Myc(353-439) and lane 10 contains the same plus 7 μl of the 26-29 kD slice. Lane 11 contains 1 μg of Protein A-

Myc(372-439) and lane 12 contains 0.5 μg of Protein A-Myc(372-439) plus 7 μl of the 26-29 kD slice.

Fig. 5. C2' Binding Activity Requires a 40-50 Kd Factor. A) 101 μg of CHO produced c-Myc was separated on an SDS gel. Proteins were recovered, resuspended in 100 μl, and renatured and analyzed by EMSA using the ERP3/4 probe. This probe contains the portion of the MLC enhancer that encompasses the μE2 site. EMSA samples contained 0.3 μg of the SDS gel load (lane 1), 7.5 μl of the proteins from the Myc slice (lane 2), or the 40-50 kD slice (lane 3), or 7.5 μl of both slices renatured together (lane 4). B) EMSA samples contained 0.9 μg purified baculovirus produced c-Myc (lane 5), 3 μl of protein from the 40-50 kD slice of a gel loaded with 400 μg CHO produced c-Myc (lane 6), or both renatured together (lane 7). The probe was ERP1/2. C) EMSA samples contained 10 μl (0.9 μg) of bacterially produced c-Myc fusion protein containing Myc amino acids 353-439 (lane 8), 0.47 μg of CHO produced c-Myc (lane 9), 5 μl of protein from the 40-50 kD slice of a gel loaded with 400 μg of the CHO Myc shown in lane 9 (lane 10), or 5 μl of the same 40-50 kD material renatured in the presence of either 0.9 μg of the baculovirus produced Myc shown in lane 5 (lane 11), 2 μl (0.18 μg) of the bacterially produced Myc fusion protein containing Myc amino acids 353-439 (lane 12), or 4 μl (0.36 μg) of the same bacterially produced Myc fusion protein (lane 13). The probe was ERP1/2.

Fig. 6. Antibodies to c-Myc Interact with the C1 and C2 Complexes. EMSA reactions were set up with the indicated Myc protein preparations (0.37 μg baculovirus produced c-Myc or 0.47 μg of CHO produced c-Myc). These reactions were preincubated 30 min on ice in the presence of the indicated antibody (α-Myc monoclonal 1F7 or a monoclonal directed against the lambda repressor, cI). 1 ng of SMS probe or μE2-containing probe number 7 (see Fig. 7) was added subsequently and binding and electrophoresis were as usual.

Fig. 7. Oligonucleotides Selected from Random Sequence after 8 Rounds of EMSA. Sequences were selected from oligonucleotides containing 20 base pairs of random sequence using a reiterative EMSA procedure described in materials and methods. Underlined nucleotides are from the PCR primer sites. Tables below the aligned sequences tabulate the frequency of each base in the 6 flanking positions surrounding the CACGTG motifs. Only bases next to a perfect fit of the CACGTG core were tabulated since sequences without this core were found not to function as high affinity binding sites (Fig. 8, and data not shown). Bold numbers adjacent to individual sequences indicate those oligonucleotides which were tested individually by EMSA in Fig. 8. Asterisks indicate additional sequences which were tested individually (data not shown).

Fig. 8. Selected Sites form Predicted Complexes. EMSA was carried out using either 2.8 ng of the SMS probe or equal amounts (1 ng) of probes 1-11 indicated in Fig. 7. Probes 1-11 were labeled and gel isolated in parallel and had approximately equal specific activities. Binding reactions contained either no additional protein (-), 0.37 μg of baculovirus produced c-Myc (B) or 0.47 μg of CHO produced c-Myc (C). Free probe is visible at the bottom of the gel.

Fig. 9. Off-Rate of the C1 and C2 Complexes. The standard EMSA reaction was scaled up for 11 samples containing 0.4 μg of purified baculovirus produced c-Myc per sample. Probe and competitor were $(USE)_3$. After allowing 20 min for binding 20 μl was loaded on a prerun EMSA gel as a measure of the starting amount of complex (ST) and enough cold competitor was added to the remaining sample to achieve a 250 fold molar excess over probe. Immediately upon addition of competitor the sample was gently mixed and 20 μl aliquots were loaded at the indicated times (0, 30 s, 1 min, 4 min, etc.). A control sample (C) was made up individually in which competitor was added prior to the start of binding to demonstrate complete competition. This sample was loaded at the same time as the ST sample. All samples were loaded on a

continuously running gel so that the complex in the starting lane runs ahead of the equivalent complex in lanes loaded later.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the description that follows, a number of terms used in recombinant DNA technology are extensively utilized. In order to provide a clearer and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

Oligomer of Interest. As used herein, an "oligomer of interest" refers to any of the following types of oligomeric proteins: first, Myc-containing oligomers including homo-oligomers of Myc peptides (a C1 complex), and hetero-oligomers containing at least one peptide of Myc and one peptide of a Myc "partner" (a C2 complex); second, oligomers that form in the presence of Myc-containing homo-oligomers or Myc-containing hetero-oligomers but which themselves do not contain the Myc peptide, such oligomers including non-Myc-containing homo-oligomers that associate in the presence of Myc and non-Myc-containing hetero-oligomers that associate in the presence of Myc (a C2´ complex).

Oligomer. An "oligomer" as it refers to proteins, means a protein composed of more than one peptide subunit, such as dimers, trimers, tetramers, etc. Such oligomeric protein may be a homo-oligomer, that is, composed entirely of two or more identical subunits; alternatively, such oligomeric protein may be a hetero-oligomer, that is, composed of at least two different peptides. Oligomers containing three or more peptides may contain more than one copy of a peptide.

C2´ Protein(s). As used herein, for convenience, a "C2´ protein" is a protein or peptide that is a member of the second class of the "oligomers-of-interest," that is, a protein that forms oligomers in the presence of Myc, c-Myc homo-oligomers or Myc-containing hetero-oligomers so as to bind to

a specific DNA sequence, but which does not contain a Myc peptide, such oligomers including non-Myc-containing homo-oligomers that associate in the presence of Myc and non-Myc-containing hetero-oligomers that associate in the presence of Myc.

Operably-linked. As used herein, two macromolecular elements are operably-linked when the two macromolecular elements are physically arranged such that factors which influence the activity of the first element cause the first element to induce an effect on the second element. For example, the transcription of a coding sequence which is operably-linked to a promoter element is induced by factors which "activate" the promoter's activity; transcription of a coding sequence which is operably-linked to a promoter element is inhibited by factors which "repress" the promoter's activity. Thus, a promoter region would be operably-linked to the coding sequence of a protein if transcription of the coding sequence activity was influenced by the activity of the promoter.

Response. As used herein, the term "response" is intended to refer to a change in any parameter which can be used to measure, indicate or otherwise describe c-Myc action or oligomer (homo-oligomer (C1 complex) or hetero-oligomer (C2 complex)) formation, including c-Myc dependent hetero-oligomerization (C2′ complex) formation. The response may be revealed as a physical change (such as a change in phenotype) or, it may be revealed as a molecular change (such as a change in a reaction rate or affinity constant). Detection of the response may be performed by any means appropriate. "Detecting" refers to any method by which such response may be evaluated so as to provide a meaningful indicia of whether the event has occurred.

Compound. The term "compound" is intended to refer to a chemical entity, whether in the solid, liquid, or gaseous phase. The term should be read to include synthetic compounds, natural products and macromolecular entities such as polypeptides, polynucleotides, or lipids,

and also small entities such as neurotransmitters, ligands, hormones or elemental compounds.

Bioactive Compound. The term "bioactive compound" is intended to refer to any compound which induces a detectable or measurable response in the methods of the invention.

Promoter. A "promoter" is a DNA sequence located proximal to the start of transcription at the 5' end of the transcribed sequence. The promoter may contain multiple regulatory elements which interact in modulating transcription of the operably-linked gene.

Expression. Expression is the process by which the information encoded within a gene is revealed. If the gene encodes a protein, expression involves transcription of the DNA into mRNA, the processing of mRNA (if necessary) into a mature mRNA product, and translation of the mature mRNA into protein.

A nucleic acid molecule, such as a DNA or gene is said to be "capable of expressing" a polypeptide if the DNA contains the coding sequences for the polypeptide and expression control sequences which, in the appropriate host environment, provide the ability to transcribe, process and translate the genetic information contained in the DNA into a protein product, and if such expression control sequences are operably-linked to the nucleotide sequence which encodes the polypeptide.

Cloning vehicle. A "cloning vehicle" is any molecular entity that is capable of delivering a nucleic acid sequence into a host cell for cloning purposes. Examples of cloning vehicles include plasmids or phage genomes. A plasmid that can replicate autonomously in the host cell is especially desired. Alternatively, a nucleic acid molecule that can insert into the host cell's chromosomal DNA is especially useful.

Cloning vehicles are often characterized by one or a small number of endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of

the vehicle, and into which DNA may be spliced in order to bring about its replication and cloning.

The cloning vehicle may further contain a marker suitable for use in the identification of cells transformed with the cloning vehicle. Markers, for example, are tetracycline resistance or ampicillin resistance. The word "vector" is sometimes used for "cloning vehicle."

Expression vehicle. An "expression vehicle" is a vehicle or vector similar to a cloning vehicle but is especially designed to provide sequences capable of expressing the cloned gene after transformation into a host.

In an expression vehicle, the gene to be cloned is usually operably-linked to certain control sequences such as promoter sequences. Expression control sequences will vary depending on whether the vector is designed to express the operably-linked gene in a prokaryotic or eukaryotic host and may additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

Host. By "host" is meant any organism that is the recipient of a cloning or expression vehicle.

a. Isolation of c-Myc Protein From Mammalian Cells and Preparation of Fractions Containing C2 and C2´ Complex Binding Activity

Although there have been previous reports of purified Myc protein, the present inventors found that the Mvc protein preparations described therein, and the methods used to isolate that protein, failed to achieve the requisite amount of yield needed to sequence characterize Myc action in mammalian sources. The inventors have overcome this problem and describe, for the first time, a unique and useful method for the isolation of highly purified mammalian c-Myc protein which provides the requisite high degree of quantity of mammalian c-Myc protein needed for the

characterization of c-Myc directed DNa binding and biological action. The inventors have also been able to purify large quantities of Myc from a recombinant insect cell system. The purified Myc protein of the invention exhibits the only known biochemical activity of c-Myc, an ability to bind the sequence CACGTG. As a direct result of the method of the invention for the isolation of c-Myc protein, the inventors were able to identify peptides that naturally associate with c-Myc in a hetero-oligomers, or peptides that naturally associate with each other as a result of the action of c-Myc, such peptides found to be present in certain column chromatography fractions of the c-Myc purification scheme.

Accordingly to the invention, purification of Myc from a mammalian source is preferably achieved utilizing a mammalian cell line that overexpresses either recombinant or non-recombinant c-Myc and is performed completely on ice or equivalent temperatures of 0-5°C, using reagents and buffers at the same temperature. For example, the overexpressing Chinese hamster ovary (CHO) cell line 5A is useful for such purification. In CHO 5A cells, recombinant mouse c-Myc is under the control of a regulatable promoter, and has been integrated and amplified in the genome of the parent CHO cell line for maximum stability and production. Except where otherwise noted, for the methods and assays of the invention, the native or recombinant Myc should include at least the two coding exons of Myc.

After collecting the cells by centrifugation using techniques known in the art, and prior to lysis of the outer cell membrane, the cells should be washed at least once in a low salt neutral buffer such as 0.9% NaCl in 10-50 mM phosphate, pH 7.0-7.5 (phosphate buffered saline, PBS) to remove remaining growth medium.

Lysis of the washed cells is also achieved in a low salt, neutral to mildly acidic lysis buffer, preferably about pH 6.8, containing at least one protease inhibitor, such as aprotinin or phenylmethylsulfonyl fluoride (PMSF), preferably containing a combination of such inhibitors. Salts such

as potassium (in the KCl form) and magnesium (in the $MgCl_2$ form) are also preferably added. In addition, nonionic detergents such as NP40 (0.5% v/v) and Na-deoxycholate (0.1 %) should be added.

Cell outer membrane lysis should be performed under conditions that lyse the host cell without lysing the nucleus, or induce significant leakage from the nuclear membrane. The cells may be allowed to sit for a short period of time, for example, 10 minutes, in the detergent-containing lysis buffer before mechanical intervention is utilized in the lysis step. Mechanical intervention is best performed with a gentle disruption of the detergent treated cells, for example, utilizing 40 strokes in a Dounce homogenizer with a type A pestle, or the equivalent of such treatment.

Nuclei may be collected from the lysed cell preparation using techniques known in the art, such as, for example, centrifugation at 1000xg for 5 min at 4°C and washed at least once in the same low salt lysis buffer used to lyse the outer cell membrane.

Nuclei are then resuspended in the low salt lysis buffer that additionally contains sufficient DNAse I and incubated for a time sufficient to efficaciously degrade the DNA in such nuclei to a size and viscosity that allows subsequent purification of the c-Myc from this preparation as described below.

Following the DNAse I treatment, the sample is diluted with a high salt neutral buffer that brings the salt (as NaCl) concentration of the sample to at least 2 M. Such high salt buffer preferably additionally also contains amounts $MgCl_2$ sufficient to maintain the same concentration of this salt in the final diluted preparation, and also additional detergent NP40so as to retain efficacious levels after sample dilution.

In mammalian host cells, c-Myc is generally tightly associated with the nuclei. Accordingly, it is necessary to solubilize c-Myc in a manner that does not destroy its biological activity or its ability to renature into a biologically active form. The residual nuclear material is first removed by

centrifugation and then the pellet resuspended for solubilization of the c-Myc. Solubilization of the c-Myc protein in a manner that destroys this association may be achieved with either sodium dodecyl sulfate (SDS) or urea at concentrations greater than 4 M. Preferably, 5M urea is utilized. Residual non-lysed nuclei may also be solubilized at this time by vigorous stirring for about 30 min. The solution is then centrifuged to pellet any remaining insoluble material prior to the subsequent chromatography steps, for example, at 5000xg for about 10 min.

The supernatant fraction recovered from the centrifugation step is applied to a DEAE Sepharose CL-6B column equilibrated in the urea-containing buffer as described above, and the column thoroughly washed with such buffer to remove unbound protein. A second wash was performed with the addition of an intermediate amount of NaCl, 0.1M NaCl to the buffer. Finally, Myc protein was eluted by raising the salt concentration in the buffer to 0.35M.

All protein eluting with the 0.35M salt wash were collected and applied to a FPLC Mono-Q column. The column was washed and with a gradient of 0.10 M NaCl to 0.35 M NaCl, followed by a 2 M NaCl step wash. Holding the gradient at intermediate salt concentrations, for example at about 0.19 M NaCl, until the end tail of the contaminating protein is finished eluting will enhance the purity of the subsequently eluted Myc protein.

Myc may be identified in the column eluent by any technique that specifically recognizes Myc protein or its activity. For example, a monoclonal antibody such as 1F7 may be used in an immunoassay for the presence of Myc protein. Alternatively, DNA binding activity to an oligonucleotide containing the sequence 5'-CACGTG-3' may be used to monitor the purification. Monoclonal antibody 1F7 is directed against the peptide sequence of amino acids 305-317 in murine c-Myc. Other Myc monoclonal antibodies useful in such assays are commercially available.

Pools of fractions from this column contained the C2 and C2´ binding activities described below, and the presence of peptides capable of entering into C2 and C2´ hetero-oligomers, and especially C2 and C2´ hetero-oligomers, may be assayed by the ability of such hetero-oligomers to bind to the DNA sequences 5'-CACGTG-3' and 5'-CAGCTG-3', respectively. Myc purified from the CHO cells appeared as multiple bands by immunoblot.

b. Purification of c-Myc and Its Partners From a Baculovirus Source

Human c-Myc may also been purified using the baculovirus overexpression system. For purification, Sf9 cells that had been infected with recombinant baculovirus carrying the c-Myc gene, using techniques known in the art were harvested just prior to the onset of lysis (−48 hours post infection). Solubilization and purification of the recombinant c-Myc were carried out as with the CHO produced Myc resulting in a yield of 2.5 mg/8x10^8 cells. Myc purified from these insect cells was apparently homogeneous by silver staining, and ran on electrophoresis as a single diffuse band of −60kD. This was in contrast to the multiple bands observed with mammalian Myc by immunoblot (Fig. 1B).

c. Detection of Sequence Specific DNA Binding Activity

The above preparations contain two sequence specific DNA-binding activities that both contain Myc protein. The first activity contains only Myc (i.e., forms the Myc homo-oligomer) and binds very weakly to sequences with the core CACGTG. The binding is assayed by determining the off rate and by competitor assays, both techniques known in the art. The binding of c-Myc homo-oligomers is characterized by an immeasurably fast off rate and by the observation that it is almost impossible to add

enough cold competitor sequence to completely compete away this complex in electrophoretic mobility shift assays (EMSA). This latter observation implies that it may not be possible to raise oligonucleotide concentrations above the K_D, thus preventing the determination of exactly what fraction of the final Myc preparations are active for sequence specific binding by the Myc homo-oligomers.

A binding site selection procedure may be used to determine the optimal binding site for Myc. Sites may be selected from a pool of random oligomers, such as 20-mers, in order to decrease bias in determining an optimal binding site. A 12 base consensus sequence of GACCACGTGCTC [SEQ ID No. 1] may be used, with the central E box core of CACGTG appearing to be most conserved. Halazonetis and Kandil (Halazonetis and Kandil, *Proc. Natl. Acad. Sci. USA* 88:6162-6166 (1991)) assumed that the flanking sequences might be symmetric, and reported an optimal sequence of GACCACGTGGTC [SEQ ID No. 2]. This sequence is quite similar to the consensus that is preferred here, differing in only the 10th position (where predominantly a C was utilized in the invention, although G is significantly represented Fig. 7, Group I). Accordingly to the invention, it is possible to select a 12 base consensus sequence from a pool of predicted complexity of 4^{20} (-10^{12}) thus indicating that Myc has a strong sequence preference despite its apparent weak binding affinity.

The second Myc containing DNA-binding complex provided in the preparations of the invention also binds to sequences with a core of CACGTG, but binds significantly more tightly than Myc alone. This complex (the C2 complex) requires a 26-29 kD factor in addition to Myc. This additional factor copurified with Myc, presumably because of similar chromatographic properties and not via association with Myc since the chromatography performed in 5M urea would denature such association. This additional factor resembles Max, a protein whose gene was recently isolated from mammalian cells, in that it does not bind efficiently to DNA by itself but can hetero-oligomerize with Myc to bind tightly to the

sequence CACGTG. However, that the factor of the invention differs from Max in its apparent size (Max is reported to migrate at 21 kD). Additionally, the Myc/Max hetero-oligomer appears to migrate at least as slowly as a Myc only complex in EMSAs, while the C2 complex of the invention migrates more rapidly than Myc alone.

In addition to the 26-29 kD factor, a second copurifying factor of 40-50 kD has been identified. The sites selected by complexes containing this factor (herein termed C2′ complexes) contained a CAGCTG core (the μE2 sequence motif) as well as flanking sequences which bear a striking resemblance to a recently reported binding site for myogenin homo-oligomers (Wright *et al.*, *Mol. Cell. Biol.* *11*:4104-4110 (1991)). Myogenin is an HLH containing protein of predicted molecular weight 32.5 kD whose optimal binding site is AACAGT/CTGTT [SEQ ID No. 3]. None of the sites (0/36) selected by the C2 or C2′ complexes of the invention contained a CAGTTG motif while roughly half of the myogenin selected sites contained such core sequences.

d. Assay for a Compound that Inhibits Myc Action

For the ease in describing these assays, C1 complex association and/or DNA binding, C2 complex association and/or DNA binding, and C2′ complex association and/or DNA binding are all referred to as c-Myc activity.

Assays for c-Myc activity may be performed *in vitro* or *in vivo*. *In vitro* assays may be performed as described in the Examples, for example, by evaluating the effects the desired compound or various amounts of such compound on the results of the electrophoretic mobility shift assay and site selection techniques that will reveal whether binding of the oligomer of interest to a specific DNA sequence motif has occurred in the presence of the compound.

For the *in vivo* assay of a compound that inhibits the desired Myc activity at least two genetic constructs are utilized. First is required a recombinant construct capable of expressing Myc is required; second is required a reporter gene whose expression is operably linked to the Myc activity and especially to the binding of the desired oligomer to the specific DNA sequence or motif.

If desired, a recombinant construct capable of expressing a C2 complex protein or C2´ complex protein may also be used. Alternatively, a host may be chosen may be chosen that naturally expresses such protein.

Recombinant constructs that are capable of expressing Myc protein may be constructed utilizing the guidelines as described below or purchased commercially.

The desired DNA binding sequence may be operably linked to any gene which confers a selectable marker in the host system. In a preferred embodiment, a marker gene which allows phenotypic selection in yeast, and especially in *Saccharomyces cerevisiae* is used.

Yeast that have been co-transformed with both an expressible *myc* gene and with the desired DNA binding sequence may be used to (1) identify the presence or absence of endogenous host proteins that interact with Myc in a C2 or C2´complex (2) classify a protein as a C1 complex protein or as a C2´ complex protein; and (3) identify and classify compounds as agents which disrupt such Myc activity. C2 complex proteins have previously also been termed Myc "partner" proteins.

All three applications are based on the same principle: in the presence of c-Myc biological activity, one of three things will happen: C1 complexes will form; C2 complexes will form; or, C2´ complexes will form. The protein complexes so formed, and especially the oligomeric complexes, will bind to a specific DNA motif, binding to such motif will be operably linked to the marker gene, and expression of the marker gene will be altered, preferably stimulated, in response to such DNA binding.

In the absence of such oligomerization, oligomer-DNA complex formation will not occur and expression of the marker protein will not be altered.

In the assays of the invention, there may be some level of binding to a desired DNA binding sequence even in the absence of c-Myc. However, when c-Myc is present in the cell, the amount and strength of the specific DNA binding is increased.

Hosts that have been co-transformed with both an expressible c-*Myc* gene and with the desired DNA binding sequence may be used to assay for the presence or absence of endogenous host proteins that interact with c-Myc activity. If such analyses reveal that the host contains c-Myc binding proteins, or c-Myc dependent oligomers which, in the presence of c-Myc specifically bind to a desired DNA sequence, such c-Myc partner protein or dependent-oligomer protein may be isolated using techniques known in the art such as gel mobility shift analysis, cDNA expression cloning vectors such as, for example, λgt10 and λgt11, or other cloning systems specifically designed for high-efficiency cloning and expression of full-length cDNA in yeast such as, for example, pG1 and pTRP56, all of which are commercially available (Clontech, Palo Alto, California).

It is not necessary that the host be completely deficient in C2 complex proteins (c-Myc partner proteins) or C2′ complex proteins to be useful in the method of the invention. As described below, if c-Myc is expressed at levels much greater than those found in the host, reporter gene transcription from endogenous partner proteins may be negligible, or of such low amount that it does not otherwise alter the utility of the methods of the invention.

If the c-Myc expression is transcribed with a strong promoter, and/or if the c-*Myc* expression cassette is supplied on a high copy number vector, the levels of c-Myc will be high enough to overcome a low level background and such c-Myc constructs may be used to analyze the ability of cloned c-Myc partners to influence c-Myc DNA binding. One of

ordinary skill in the art can adapt the expression system to the level of expression desired using methods known in the art.

The C2 complex protein (the partner protein), or the C2´ complex protein, if supplied as a recombinant construct to the host cell, should be capable of expressing at levels comparable to that of the c-Myc protein. C2 complex proteins may be identified by utilizing a phage plaque assay, as described in the commonly-owned, copending U.S. patent application entitled "Protein Partner Screening Assays and Uses Thereof," Application No. 510,254, filed April 19, 1990, and incorporated herein by reference. Proteins identified by such screening assay can be subcloned into eukaryotic expression vectors known in the art and commercially available so as to provide a recombinant source of partner protein gene expression.

The genetic constructs of the invention may be placed on different plasmids, or combined on one plasmid. A construct may also be inserted into the genome of a host cell. Preferably, the construct coding for the c-Myc protein and the construct coding for the C2 complex protein or the C2´ complex protein are provided to the host on two different plasmids.

It is important to establish that the effect of the compound is due to an effect on c-Myc activity and not an effect on the activity of the reporter product *per se*. Such effect can be established by comparing the results found in hosts which lack either the c-Myc expression vector or the C2 or C2´ protein expression vector or both.

The desired DNA binding motif may be located at any site in the transcription cassette of the reporter gene which allows for the transcription of that gene to be operably-linked to binding of the desired oligomer. Thus, such motif may be located 5' to the transcriptional start site or 3' to the transcriptional start site, for example, in an intron, similar to its location relative to the promoter region in the immunoglobulin genes.

The reporter gene whose expression is operably linked to c-Myc activity and especially to oligomer DNA binding may be any gene whose expression can be monitored. Any detectable phenotype change may serve

as the basis for the methods of the invention. In a preferred embodiment, the reporter gene is a gene not normally expressed by the host, or a gene that replaces the host's endogenous gene. Any reporter gene which is capable of being operably-linked to a promoter capable of responding to the binding of the oligomer of interest to the specific target DNA sequence may be used.

Especially, for example, genes that endow the host with an ability to grow on a selective medium are useful. For example, in yeast, use of the yeast LEU2 gene as a reporter gene in strains that normally lack LEU2 allows such yeast to grow on leucine as a sole carbon source. Expression the reporter gene is monitored by merely observing whether the host possesses the ability to grow on leucine. In a similar manner, use of the *suc*2 gene as a reporter gene would allow growth of the a suc2$^-$ yeast host on sucrose to be used as the detection method. In both examples, growth on the indicated substrate could be used to indicate specific DNA binding of the oligomer of interest and lack of such growth could be used to indicate lack of binding or lack of oligomer formation.

In another example, a construct (and host) which is *gal*1$^+$*gal*10$^-$ would respond to galactose in the medium; a construct (and host) which is *lac*2$^+$*gal*1$^+$ would be lactose sensitive. Other reporter genes include *his*3, *ura*3 and *trp*5. One of ordinary skill in the art can imagine many other appropriate reporter systems which would reveal the presence or inhibition of DNa binding or biological activity of the oligomer of interest.

Reporter constructs in which the desired DNA sequence motif and the *lac*Z reporter gene are operably linked will express β-galactosidase in response to binding of a c-Myc activity induced oligomer binding to such DNA sequence. Such expression can be easily scored by monitoring the ability of the host to produce β-galactosidase (Maniatis, T. *et al.*, *Molecular Cloning (A Laboratory Manual)*, 2nd edition, Cold Spring Harbor Laboratory, 1989). The production of β-galactosidase may be visually monitored by detecting its activity to reduce the chromophoric dye,

X-gal (commercially available from International Biotechnologies, Inc., New Haven, CT). β-galactosidase reduces X-gal to a form which possesses a blue color. In another embodiment, the coding sequence of chloramphenicol acetyltransferase (CAT) is used as the reporter gene.

Any detection method that can identify expression of the reporter gene may be used. For example, levels of the product of the reporter gene may be directly assayed with an immunoassay. Such immunoassays include those wherein the antibody is in a liquid phase or bound to a solid phase carrier. In addition, the reporter gene can be detectably labeled in various ways for use in immunoassays. The preferred immunoassays for detecting a reporter protein using the include radioimmunoassays, enzyme-linked immunosorbent assays (ELISA), or other assays known in the art, such as immunofluorescent assays, chemiluminescent assays, or bioluminescent assays.

In an assay to screen for the ability of a compound to alter binding of the oligomer of interest, yeast strains that express such the desired peptide or peptides and which contain the related DNA binding sequence motif, may be plated and grown as lawns and the compound to be tested may be applied to the plates on a filter paper disk that is impregnated with such compound. Alternatively, the compound may be incorporated into the media within which the host cells are growing.

One may be able to detect the ability of a compound to alter c-Myc activity by the appearance of a zone, which often resembles a halo, around the compound-impregnated disk. If for example, the compound is toxic to the host's survival *per se*, the host will not grow in the zone containing the compound.

The methods of the invention can be used to screen compounds in their pure form, at a variety of concentrations, and also in their impure form. The methods of the invention can also be used to identify the presence of such inhibitors in crude extracts, and to follow the purification of the inhibitors therefrom. The methods of the invention are also useful in

the evaluation of the stability of the inhibitors identified as above, to evaluate the efficacy of various preparations.

The permeability of cells to various compounds can be enhanced, if necessary, by use of a mutant cell strain which possess an enhanced permeability or by using compounds which are known to increase permeability. For example, in yeast compounds such as polymyxin B nonapeptide may be used to increase the yeast's permeability to small organic compounds. In cells from the higher eukaryotes, dimethyl sulfoxide (DMSO) may be used to increase permeability. Analogs of such compounds which are more permeable across yeast membranes may also be used. For example, dibutyryl derivatives often display an enhanced permeability.

In a preferred embodiment, the genetic constructs and the methods for using them are utilized in eukaryotic hosts, and especially in yeast, insect and mammalian cells. The introduced sequence is incorporated into a plasmid or vector capable of either autonomous replication or integrative activity.

The sequence of c-Myc is known (Battey, J. *et al.*, *Cell 34*:779-787 (1983)) and probes which are capable of identifying a c-*Myc* clone are commercially available (New England Nuclear/DuPont Biotechnology Boston, MA).

The DNA sequence of the desired gene may be chemically constructed if it is not desired to utilize a clone of the genome or mRNA as the source of the genetic information. Methods of chemically synthesizing DNA are well known in the art (*Oligonucleotide Synthesis, A Practical Approach*, M.J. Gail, ed., IRL Press, Washington, D.C., 1094; *Synthesis and Applications of DNA and RNA*, S.A. Narang, ed., Academic Press, San Diego, CA, 1987). Because the genetic code is degenerate, more than one codon may be used to construct the DNA sequence encoding a particular amino acid (Watson, J.D., In: *Molecular Biology of the Gene*, 3rd edition, W.A. Benjamin, Inc., Menlo Park, CA, 1977, pp. 356-357).

To express the recombinant constructs of the invention, transcriptional and translational signals recognizable by the host are necessary. A cloned protein encoding DNA sequence, obtained through the methods described above, (preferably in a double-stranded form), may be operably-linked to sequences controlling transcriptional expression in an expression vector, and introduced, for example by transformation, into a host cell to produce recombinant proteins useful in the methods of the invention, or functional derivatives thereof. Such techniques are well known in the art (*Recombinant DNA Methodology*, Wu, R. *et al.*, eds., Academic Press, (1989); Maniatis, T. *et al.*, *Molecular Cloning (A Laboratory Manual)*, second edition, Cold Spring Harbor Laboratory, 1989).

Transcriptional initiation regulatory signals can be selected which allow for repression or activation of the expression of the c-Myc construct or construct of the recombinant C2 complex peptide (or the C2′ peptide), or both, so that expression of such constructs can be modulated, if desired. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical regulation, for example, by a metabolite, salt, or substrate added to the growth medium.

Where the native expression control sequences signals do not function satisfactorily in the host cell, then sequences functional in the host cell may be substituted.

Expression of the constructs of the invention in different hosts may result in different post-translational modifications which may alter the properties of the proteins expressed by these constructs. It is necessary to express the proteins in a host wherein the ability of the protein to retain its biological function is not hindered. Expression of proteins in yeast hosts is preferably achieved using yeast regulatory signals. The vectors of the invention may contain operably-linked regulatory elements such as

upstream activator sequences in yeast, or DNA elements which confer species, tissue or cell-type specific expression on an operably-linked gene.

In general, expression vectors containing transcriptional regulatory sequences, such as promoter sequences and transcription termination sequences, are used in connection with a host. These sequences facilitate the efficient transcription of the gene fragment operably-linked to them. In addition, expression vectors also typically contain discrete DNA elements such as, for example, (a) an origin of replication which allows for autonomous replication of the vector, or, elements which promote insertion of the vector into the host's chromosome in a stable manner, and (b) specific genes which are capable of providing phenotypic selection in transformed cells. Eukaryotic expression vectors may also contain elements which allow it to be maintained in prokaryotic hosts; such vector are known as shuttle vectors.

The precise nature of the regulatory regions needed for gene expression will vary between species or cell types and there are many appropriate expression vector systems that are commercially available.

In a highly preferred embodiment, yeast are used as the host cells. The elements necessary for transcriptional expression of a gene in yeast have been recently reviewed (Struhl, K. *Ann. Rev. Biochem. 58*:1051-1077 (1989)). In yeast, most promoters contain three basic DNA elements: (1) an upstream activator sequence (UAS); (2) a TATA element; and, (3) an initiation (I) element. Some promoters also contain operator elements. Methods in yeast genetics are well known (Struhl, K. *Nature 305*:391-397 (1983); Sherman, *et al., Methods in Yeast Genetics*, Cold Spring Harbor Laboratory (1983)).

In another embodiment, mammalian cells are used as the host cells. A wide variety of transcriptional and translational regulatory signals can be derived for expression of proteins in mammalian cells and especially from the genomic sequences of viruses which infect eukaryotic cells.

Once the vector or DNA sequence containing the construct(s) is prepared for expression, the DNA construct(s) is introduced into an appropriate host cell by any of a variety of suitable means, for example by transformation. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene sequence(s) results in the production of the protein. This expression can take place in a continuous manner in the transformed cells, or in a controlled manner.

Genetically stable transformants may be constructed with episomal vector systems, or with integrated vector systems whereby the fusion protein DNA is integrated into the host chromosome. Such integration may occur *de novo* within the cell or be assisted by transformation with a vector which functionally inserts itself into the host chromosome, for example, with retroviral vectors, transposons or other DNA elements which promote integration of DNA sequences in chromosomes.

Cells which have been transformed with the DNA vectors of the invention are selected by also introducing one or more markers which allow for selection of host cells which contain the vector, for example, the marker may provide biocide resistance, e.g., resistance to antibiotics, or heavy metals, such as copper, or the like.

The transformed host cell can be fermented or cultured according to means known in the art to achieve optimal cell growth, and also to achieve optimal expression of the cloned protein sequence fragments. As described hereinbelow, a high level of recombinant protein expression for the cloned sequences coding for the proteins can be achieved according to a preferred procedure of this invention.

The methods of the invention are not intended to be limited to c-Myc and possess utility for the characterization of inhibitors against any Myc protein, such as, for example, N-Myc and L-Myc. The C2 complex peptides of the invention may interact with more than one Myc protein and

the C2′ complex peptides of the inventions may form as the result of the activity of more than one Myc protein.

The following examples further describe the materials and methods used in carrying out the invention. The examples are not intended to limit the invention in any manner.

EXAMPLES

Example 1

Materials and Methods

Cell Growth and Myc Overexpression: The 5A cell line was maintained in spinner culture under selection with 80 μM methotrexate. Protein purification started with roughly 6 liters of cells at $8x10^5$/ml grown up without selection. Heat shock promoter induction was achieved by resuspension in preheated fresh media (43°C) at 1/3 the original volume. Cells were incubated with stirring at 43°C for 1 h. To allow translation of the accumulated mRNA, cells were transferred to 37°C culture conditions for 3 h. Cells were then subjected to the purification described below.

The baculovirus overexpression vector was constructed by insertion of the BamH1/BclI fragment of pGEMMycB [Halazonetis and Kandil, *Proc. Natl. Acad. Sci. USA* 88:6162-6166 (1991)] into the BamH1 site of a baculovirus expression vector, pVL941, obtained from the laboratory of Max Summers (Texas A&M University, College Station, Texas). The resulting plasmid contained the entire coding sequence of the human Myc gene including 6 nucleotides 5′ of the initiation codon and 3′ untranslated sequence extending to the genomic RsaI site. Sf9 cells were grown and infected with recombinant baculovirus according to the methods of Summers [Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experiment Station Bulletin No. 1555] with minor changes. Cells were passaged in

spinner culture and plated on 150 mm diameter tissue culture plates for protein production. Cells were infected and harvested approximately 48 h post infection by scraping. Cells were then washed in PBS and subjected to the purification described below.

The Protein A-c-Myc fusion protein was expressed in the *E. coli* AR68 strain from a previously published pRIT2T vector [Dang, C.V., *Anal. Biochem. 174*:313-317 (1988)] which fused the Ig binding portion of protein A to either amino acids 353-439 or amino acids 372-439 of c-Myc. Growth and induction of the cells was as per Dang *et al.* [*Anal. Biochem. 174*:313-317 (1988)].

Protein Purification: All purification steps were carried out on ice or with ice cold buffers unless otherwise stated. Cells may be used fresh or stored quick frozen in liquid nitrogen for larger batch preparations. 5A or Sf9 cells were washed in phosphate-buffered saline (PBS) and resuspended at $2.1x10^7$ cells/ml in Low Salt Lysis Buffer (20 mM HEPES pH 6.8, 5 mM KCl, 5 mM $MgCl_2$, 0.5% NP40, 0.1 % Na-deoxycholate, 1 μg/ml aprotinin, and 0.1 mM PMSF) [Evan and Hancock, *Cell 43*:253-261 (1985)]. After 10 min cells were subjected to 40 strokes in a Dounce homogenizer with a type A pestle. Nuclei were pelleted at 1000xg, 5 min, 4°C, washed once in 50 ml Low Salt Lysis Buffer, resuspended at $2.5x10^8$ nuclei/ml in Low Salt Lysis Buffer containing 50 μg/ml DNAse I and incubated at 4°C for 1 h. An equal volume of ice cold 2X High Salt Buffer (2x concentrations: 20 mM Tris, pH 7.4, 4 M NaCl, 1 mM $MgCl_2$, and 0.1% NP40) [Evan and Hancock, *Cell 43*:253-261 (1985)] was then added, mixed gently, and incubated for 10 min. The residual nuclear material (including the c-Myc protein) was pelleted (2000xg, 10 min, 4°C) and resuspended for solubilization at $5.5x10^7$ nucleus equivalents/ml in Buffer A (50 mM Tris, pH 8.0, 2 mM EDTA, 5 % glycerol, .1 mM DTT, and .1 mM PMSF) [Watt *et al.*, *Mol. Cell. Biol. 5*:448-456 (1985)] containing 5 M urea (referred to as 5 M urea Buffer A) achieved by

dilution of a freshly deionized stock of 6 M urea. This and all buffers used on columns were passed through 0.2 pore μm filter units. Residual nuclei were solubilized by vigorous stirring on ice for 30 min. This protein solution was centrifuged (10 min, 5000xg, 4.C) to pellet any insoluble material prior to chromatography. The supernatant was loaded on a 10 ml DEAE Sepharose CL-6B (Pharmacia) column equilibrated with 5 column volumes of 5 M urea Buffer A. Sample loading was at 0.1 ml/min and column washing and elution were at 0.4 ml/min. After loading, the column was washed with 3 volumes 5 M urea Buffer A containing no additional salt followed by 4 volumes of the same buffer containing 0.1 M NaCl. Myc protein was eluted in the following elution step at 0.35 M NaCl. The protein containing fractions of this 0.35 M NaCl step were pooled and diluted with fresh 5 M urea Buffer A to 0.1 M NaCl and loaded onto a 1 ml FPLC Mono-Q column (Pharmacia) run at 0.5 ml/min. The Mono-Q column was eluted with a programmed gradient of 5 ml spanning 0.10 M NaCl to 0.35 M NaCl followed by a 2 M NaCl step. For enhanced purity the gradient was held manually at approximately 0.19 M until the major contaminating protein finished eluting as determined by an in line UV monitor. In the initial development of the purification protocol fractions from the columns were assayed for Myc by slot blotting followed by visualization using the 1F7 monoclonal antibody and ^{125}I-labeled secondary antibody. For later preparations silver staining of SDS-PAGE allowed sufficiently unambiguous identification of the Myc proteins and provided an assessment of the purity of given fractions. The Myc containing fractions were pooled based on purity and dialyzed against buffer containing 20 mM Tris, pH 7.8, 50 mM KCl, 10 % glycerol, 0.1 mM DTT, and 0.1 mM PMSF (referred to as Dialysis Buffer) in bags of SpectroPor 2 membrane for 3 changes, 2 liters each, for a minimum of 3 h each. Pools of fractions prepared this way contained C1 and C2 (and C2') binding activities. To obtain pure C1 binding activity the Myc-containing

Mono Q fractions were assayed by EMSA and those free of C2 binding activity were pooled and dialyzed separately.

The bacterially produced Protein A-c-Myc fusion protein was partially purified by differential centrifugation and solubilized in 5 M urea according to Watt *et al.* [Bagchi *et al.*, *Mol. Cell. Biol.* 7:4151-4158 (1987)] with the following minor modifications: Protease inhibitors were present in the initial lysis buffer (10 μg/ml pepstatin, 1 mM PMSF, 50 μg/ml aprotinin, 2 μg/ml leupeptin, 10 mM Na-metabisulfite, and 1 mM benzamidine) and cells were sheared by 6 bursts of 15 s each in a Cuisinart MiniMate on ice. The urea solubilized material was cleared of insoluble material by centrifugation (10,000xg, 10 min, 4°C) and dialyzed into Dialysis Buffer containing 0.5 mM DTT. Precipitated material was removed by centrifugation (15,000xg, 20 min, 4°C). Protein A-Myc fusion protein was purified from the supernatant by IgG affinity essentially according to Nilsson *et al.* [*EMBO J.* 4:1075-1080 (1985)]. A 1 ml aliquot of supernatant was incubated with 0.1 ml of a 50% slurry of IgG Sepharose 6 fast flow (Pharmacia) rocking for 1 h at 4°C. The pellet was washed twice with Buffer A and the fusion protein eluted with 0.3 M lithium diiodosalicylate (LIS). The eluate was then dialyzed extensively to remove the LIS (initially against Buffer A at room temperature to avoid LIS precipitation, then against Dialysis Buffer 4°C). The two bacterially expressed Myc preparations were compared by Coomassie staining of SDS-PAGE to ensure that equal amounts of the fusion proteins were used for experiments.

N-Terminal Sequencing: The 3 bands of purified Myc from 5A cells were individually isolated by electroelution according to Hunkapiller *et al.* [*Meth. Enz.* 91:227-236 (1983)]. Preparative SDS-PAGE was carried out and protein bands excised after visualization with Coomassie Brilliant Blue R-250. After electroelution the material was precipitated 2 times with

methanol/acetone and submitted for N-terminal sequencing by Edman degradation.

Antibodies: The monoclonal antibody, 1F7 (a generous gift of R. Chizzonite, Hoffman LaRoche), is directed against the peptide sequence comprising amino acids 305-317 in murine c-Myc [Miyamoto *et al.*, *Proc. Natl. Acad. Sci. USA* 82:7232-7236 (1985)]. The antibody directed against cI was monoclonal 51F [Breyer and Sauer, *J. Biol. Chem.* 264:13348-13354 (1989)] which had been purified by ammonium sulfate precipitation and chromatography on QAE Sephadex.

Electrophoretic Mobility Shift Assay (EMSA): Radiolabeled probes were produced via a Klenow fill in of annealed oligonucleotides containing 4 base 5' overhangs at each end (see table below for sequences). Binding reactions took place in a final volume of 20 μl containing 2 ng of labeled probe, 125 ng poly d(IC), an indicated amount of protein, and the following final buffer conditions: 10 mM Tris, pH 7.5, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 1 mM $MgCl_2$ and 5% glycerol. Binding reactions were allowed to proceed for 20 min at room temperature and were then loaded immediately on a 4% polyacrylamide gel which had been prerun at least 1 h at 10V/cm. Electrophoresis was for 1.5 h at 10V/cm in 0.5x TBE.

Cut and Renature: The method of Bagchi *et al.* [Bagchi *et al.*, *Mol. Cell. Biol.* 7:4151-4158 (1987)] was followed except for the final dialysis step. Precipitated protein samples containing BSA as carrier protein were solubilized in 6 M guanidine-hydrochloride (200 μl unless otherwise indicated) according to Bagchi *et al.* [*Mol. Cell. Biol.* 7:4151-4158 (1987)]. Directly prior to analysis by EMSA the samples were subjected to dialysis alone or in combination with another sample in a total volume of 15 μl.

Equal volumes of each sample were used in a given experiment and the volume was brought to 15 μl using 6 M GuHCl containing 0.1 mg/ml BSA. Dialysis was against 40 ml of Dialysis Buffer carried out for 1 h at 4°C on floating 13 mm membrane discs (Millipore #VSWP-013, pore size 0.025 μm).

Site Selection from Random Sequences: The following procedure was devised based on the method of Pollock and Treisman [*Nucl. Acids Res.* *18*:6197-6204 (1990)]. A 52 base oligonucleotide "randomer" (see table below) was annealed to the following 16 base primer: Xho I primer 5' CCGATATCTCGAGACGG 3', [SEQ ID No. 4]. The annealed primer was extended using Klenow and nucleotides (0.2 mM cold dNTPs and 0.4 μM α^{32}P-dCTP 800Ci/mmol) to create a pool of double stranded probes representing approximately 4^{20} sequences. The initial round of binding site selection by EMSA utilized 200 ng of this pool and either 0.37 μg of baculovirus produced c-Myc or 0.5 μg of CHO produced c-Myc. Other parameters were as previously described for EMSA. Lanes containing randomer probes were alternated with reference lanes containing 2 ng $(USE)_3$ probe and 0.37 μg of baculovirus c-Myc. The completed EMSA gel was electroblotted onto NA45 membrane (200 mA, 2.5 hrs) and the wet membrane was wrapped in plastic wrap and exposed for at least 1.5 hrs. The regions of the randomer lanes corresponding to the visible C1 and C2 complexes of the reference lanes were excised and eluted with 100 of elution solution (10 mM Tris, pH 8.0, 1 mM EDTA, 1 M NaCl) 30 min at 68°C. The liquid was transferred to a fresh tube and the membrane was rinsed with 100 μl TE which was added to this eluate. After pelleting the particulate debris, the DNA was precipitated with the addition of 10 μg glycogen, 2 μl 1 M $MgCl_2$ and 2.5 volumes of ethanol. The pellet was rinsed with 70% ethanol, dried, and the recovery assessed by scintillation counter. The entire pellet of each sample (~29-57 pg) was resuspended in 10 μl 10x PCR buffer (500 mM KCl, 100 mM Tris, pH 8.4, 1 mg/ml

gelatin, 15 mM $MgCl_2$) and 32 μl water. After addition of 1 μl each of 100 μM Xho I primer and Xba I primer (5' GGACGATCTAGATTCG 3', [SEQ ID No. 5]), 5 μl of nucleotide mix (2 mM dNTPs and 4 μM α^{32}P-dCTP 800Ci/mmol), and 1 U Taq polymerase the reactions were overlaid with paraffin oil and subjected to 20 cycles of PCR in an Ericomp machine: 2 min 94°C, 20x (15 sec 95°C, 15 sec 55°C), 10 min 72°C. The products were gel purified on 10% acrylamide and precipitated using 10 μg glycogen as carrier. Recovery was measured by scintillation counter and after resuspension in the EMSA reaction buffer (10 mM Tris, pH 7.5, 50 mM KCl, 1 mM EDTA, 1 mM $MgCl_2$, and 5% glycerol) this probe was used for the next round of EMSA selection. Subsequent cycles were primarily as above, however, 50 ng of probe was used. Eight rounds of selection and amplification were completed for the baculovirus c-Myc and seven rounds for the CHO c-Myc. After the final PCR reaction the products were extracted twice with phenol, twice with ether, and precipitated prior to digestion with Xho I and Xba I. After gel isolation the appropriate fragment was subcloned into the Bluescript SK vector (Stratagene) and sequenced by standard procedures.

Oligonucleotides Used: Oligonucleotide sequences that were used are shown below, with the E-Box core sequences underlined:

SEQ ID NO. 6:
$(\mu E2)_3$ 5' GATCTCTGCAGCAGCTGGCAGCAGCTGGCAGCAGCTGGCG 3';

SEQ ID NO: 7:
$(\mu E3)_3$ 5' GATCTGCAGTCATGTGGCGTCATGTGGCGTCATGTGGCAG 3';

SEQ ID NO: 8:
$(USE)_3$ 5' GATCTGCAGTCACGTGGCGTCACGTGGCGTCACGTGGCAG 3';

SEQ ID NO. 9:

MLC-A 5′ TCGACGTCGCAGCAGGTGCAG 3′;

SEQ ID NO. 10:
MLC-B 5′ TCGACCCCACCAGCTGGCGAG 3′;

SEQ ID NO. 11:
ERP1/2 5′ AGCTTCGAACACCTGCAGCAGCTGGCAGGAAGCAGGCCTA 3′;

SEQ ID NO. 12:
ERP3/4 5′ AGCTTTAAAATCCCCACCAGCTGGCGAAGCAACAGGTGCA 3′;

SEQ ID NO. 13:
HSE 5′ AATTGCGAAACCCCTGGAATATTCCGACCTGGCAGCCTC 3′;

SEQ ID NO. 14:
SMS 5′ TCGACTTTAGACCACGTGGTCCCCTCGA 3′;

SEQ ID NO. 15:
Randomer 5′ GGACGATCTAGATTCG$(N)_{20}$CCGTCTCGAGTATCGG 3′.

Example 2

Purification of c-Myc Protein

A primary goal of this work was to purify and characterize Myc from a mammalian source. An inducible mammalian overexpression system that has been described previously was utilized (Wurm *et al.*, *Proc. Natl. Acad. Sci. USA 83*:5414-5418 (1986)). Briefly, the two coding exons of the mouse c-Myc gene under the control of a *Drosophila* heat shock promoter had been integrated and amplified in the genome of a Chinese hamster ovary (CHO) cell line. This overexpressing cell line, 5A, was adapted to spinner culture. Heat shock (43°C) induces transcription of the amplified *myc* genes while a subsequent 2 hour recovery period at normal growth temperature (37°C) permits translation. The resulting products were

phosphoproteins of 60, 62, and 72kD which were immunoprecipitable with Myc-specific monoclonal antibodies (Wurm *et al.*, *Proc. Natl. Acad. Sci. USA 83*:5414-5418 (1986)). The c-Myc produced was tightly associated with the nuclei and attempts to solubilize it using a number of detergents, salts, and reducing agents were unsuccessful (data not shown). Significant solubilization was achieved however with either SDS or with urea at concentrations greater than 4 M. For purification, the Myc was solubilized with 5 M urea and chromatographed on DEAE resin and FPLC Mono-Q as described in materials and methods. The presence of Myc in the column fractions was assayed by immunoblot using an antipeptide monoclonal antibody, 1F7 (Miyamoto *et al.*, *Proc. Natl. Acad. Sci. USA 82*:7232-7236 (1985)). This purification procedure yielded 150μg of c-Myc per liter of spinner cells ($8x10^8$ cells). The Myc appeared to be 95% homogeneous as judged by silver staining (Fig. 1A).

An alternative translation start site for c-Myc accounts for some of the molecular weight heterogeneity of c-Myc translated *in vitro* and expressed in several cell lines (Hann *et al.*, *Cell 52*:185-195 (1988)). This alternate site is upstream from the canonical start site, however, and is not present in our overexpressor gene. N-terminal sequence analysis of each of the three prominent Myc bands described above revealed, as expected, the sequence predicted by the canonical start site (data not shown), although the N terminal methionine was not present, presumably because of N terminal processing. Therefore the potentially important differences in apparent molecular weight that are observed might be attributed to post-translational modifications and not N-terminal heterogeneity.

Human c-Myc has also been purified using the baculovirus overexpression system. For purification, Sf9 cells that had been infected with recombinant virus were harvested just prior to the onset of lysis (~48 hours post infection). Myc produced using the baculovirus system has been previously reported to be both phosphorylated and tightly associated with the nucleus (Miyamoto *et al.*, *Mol. Cell. Biol. 5*:2860-2865 (1985)).

Solubilization and purification were carried out as with the CHO produced Myc resulting in a yield of 2.5 mg/$8x10^8$ cells. Myc purified from these insect cells was apparently homogeneous by silver staining, and ran as a single diffuse band of ~60kD (Fig. 1B). This was in contrast to the multiple bands observed with mammalian Myc by immunoblot (Fig. 1B).

Discussion

Myc was purified to near homogeneity from overexpressing mammalian cells and baculovirus infected cells. The mammalian derived protein appears to be highly modified in contrast to Myc expressed in and purified from insect cells. Up to 19 distinct species of c-Myc can be identified by two dimensional gel electrophoresis (Fig. 1). These species differ both in size (approximate MRs of 60,000, 62,000 and 72,000, although this estimate of size can vary with different gel conditions) and in pI. These differences in pI might in part be attributed to differences in phosphorylation, as c-Myc is known to be phosphorylated and the change in pI of the species is consistent with incremental additions of phosphate. Although the Myc produced by the baculovirus overexpression system does not demonstrate the same molecular weight heterogeneity as the mammalian protein, it too is phosphorylated (Miyamoto *et al.*, *Mol. Cell. Biol. 5*:2860-2865 (1985)). The specific sites of phosphorylation have not been determined for either Myc preparation and other as yet unidentified modifications may distinguish these two Myc preparations.

Example 3

Specific DNA Binding Activity Present in Purified c-Myc

The presence of a B-HLH domain in c-Myc suggested that it would bind to an E-Box-like sequence of the general pattern CANNTG. These sites were first identified in immunoglobulin enhancers but have since been

found in many other tissue specific enhancers. It was first determined if any of these would be bound by the purified c-Myc proteins described in Example 2. A large number of E box related sequences were screened by electrophoretic mobility shift assays (EMSA). Those shown in Fig. 2 include synthetic oligonucleotides containing trimers of either the μE2 (CAGCTG) or μE3 (CATGTG) sites of the immunoglobulin enhancer and a trimer of the Adenovirus major late promoter upstream element (USE) (CACGTG). Two sites from the myosin light chain (MLC) enhancer are also shown: the A site (CAGGTG) which resembles the kE2 immunoglobulin enhancer site, and the B site (CAGCTG) which has the same core sequence as the μE2 site. The heat shock element (HSE) served as a control since its sequence does not resemble an E-Box core.

Three specific binding activities were detected in this assay forming complexes referred to as C1 (USE specific), C2 (USE specific), and C2′ (μE2 specific). As demonstrated below, despite the comigration of C2 and C2′, these represent separate complexes based on observed differences in protein composition as well as binding specificity. The data presented argue that the C1 complexes are formed by homo-oligomers of Myc while formation of the C2 and C2′ complexes each require an additional protein. The slowly migrating complex (C1) formed most readily on the USE (Fig. 2, lanes 5, 11, and 12), less well on the similar μE3 site (Fig. 2, lanes 2 and 8), and not at all on the other E-Box and non-E-Box sites tested. CHO and baculovirus Myc preparations were similar with regard to the C1 complex, however they differed with regard to the faster migrating complexes. In the mammalian Myc assays the C2′ complex formed on the μE2 site of the immunoglobulin enhancer and the is μE2-like sequence of the MLC-B site (Fig. 2, lanes 1 and 4). Baculovirus Myc contained no binding activity with this specificity (Fig. 2, lanes 7 and 10). In contrast, formation of the C2 complex was detected using either Myc preparation. The C2 complex formed most readily on the USE site (Fig. 2, lanes 5, 11, and 12) and less well on the similar μE3 sequence (Fig. 2, lanes 2 and 8).

Very little if any binding was detected on the kE2-like sequences (MLC-A Fig. 2, lanes 3 and 9, and μE5, data not shown). No specific binding was found on non-E-Box sequences such as the HSE (Fig.2, lane 6 and 13).

Competition experiments were performed on the three binding activities C1, C2, and C2′ to further characterize their specificity (data not shown). In experiments using μE2, μE3, USE, μE5, or HSE sequences as competitors, competition of the C2′ complex formed on the μE2 probes was most easily achieved with the μE2 oligos while the C2 complexes were preferentially competed by the USE sequence. The C1 complex was also competed most efficiently by the USE sequence. A detailed analysis of the binding specificities of these complexes is presented below.

Example 4

Proteins Responsible for Formation of C1, C2, and C2′ Complexes

One scenario suggested by the differences in binding is that Myc might not be the only protein involved in formation of the three complexes. To distinguish the role of c-Myc from other copurifying proteins in the formation of the observed complexes cut and renature experiments were performed as follows. Preparative amounts of Myc were separated by SDS-PAGE. Proteins were electroeluted from various molecular weight slices, precipitated, solubilized in guanidine-hydrochloride and dialyzed to renature for analysis by EMSA. The C1 complex binding activity may be renatured from the Myc containing slices of either baculovirus or mammalian preparations (Fig. 3) while no other slices from the entire gel contained C1 activity (data not shown). These data argue that Myc alone is the protein responsible for the C1 complex, and that full length Myc protein as expressed in eukaryotic cells can bind specifically to sites with the core sequence CACGTG.

Analysis of the proteins responsible for formation of the C2 and C2′ complexes was achieved with additional cut and renature experiments

performed as described above. EMSA using the USE probe revealed no single slice from CHO or baculovirus preparations which contained detectable C2 binding activity (data not shown). However, this activity was recovered by renaturing proteins from a 26-29 kD slice together with proteins in the 60-70 kD Myc containing slice (Fig. 4, lanes 1-8). The 26-29 kD component was present in gels loaded with either CHO or baculovirus produced c-Myc, and, when renatured with Myc, demonstrated the same specificity as the C2 complex in the loaded material. Renaturation of the 26-29 kD slice with BSA or protein A did not yield USE binding activity suggesting that Myc plays a specific role in the recovery of C2 binding activity.

To examine further the roles of copurifying proteins and of Myc modifications in the observed binding, Myc was also purified from a bacterial overexpression system. The expression system and purification method used were those of Chi Dang and colleagues (see materials and methods). The bacterially produced protein contains the IgG binding segment of protein A fused to the C-terminal 85 amino acids of Myc, the segment of Myc which contains the B-HLH and leucine zipper motifs. For many of the B-HLH proteins, the small region of the protein containing the B-HLH motif is not only necessary but fully sufficient for DNA binding if the correct oligomerization partner is present. This protein was able to form the C1 complex on the USE probes (Fig. 4, lane 9) and to combine with the 26-29 kD factor to create the C2 complex (Fig. 4, lane 10). Competition experiments confirmed the specificity of this reconstituted C2 complex. The C1 and C2 complexes formed using this bacterial fusion protein migrated more rapidly than those formed using full length c-Myc (compare Fig. 4, lane 8 with lanes 9 and 10). This may be due to the difference in size between the full length c-Myc (60-72 kD) and the protein A-Myc fusion protein (~38 kD) and therefore the mobility of the C2 complex may be interpreted as an indication that Myc is physically present

in the C2 complex, presumably as part of a hetero-oligomer with the 26-29 kD factor.

Analogous experiments were carried out using a similar bacterial fusion protein containing only the C-terminal 67 amino acids of c-Myc. This protein contains most of the HLH domain and the entire leucine zipper domain but no basic region. Although this protein is capable of forming homo-oligomers in solution (Gentz *et al.*, *Science 243*:1695-1699 (1989)), it was unable to bind to DNA to form the C1 complex and was also unable to combine with the 26-29 kD factor to create any USE binding activity (Fig. 4, lane 12). These data argue that the role of Myc in the C2 hetero-oligomer requires an intact basic region, the region responsible for specific DNA contacts in other B-HLH proteins.

Using cut and renature experiments the μE2 binding activity responsible for the C2′ complex was able to be identified. A small amount of the C2′ complex was frequently seen with proteins from the slice encompassing the 40-50 kD molecular weight range of mammalian Myc preparations (Fig. 5A). Although no C2′ complex was ever seen with the Myc containing slice alone, renaturation of the protein from the Myc slice with the 40-50 kD slice reproducibly increased the amount of C2′ complex formed. Both the baculovirus produced Myc and the bacterially expressed fusion protein containing the basic region, which do not form complexes themselves on μE2 probes, were also able to increase the amount of complex formed by the 40-50 kD slice obtained from mammalian preparations (Fig. 5B and C). Surprisingly the bacterially produced Myc lacking the basic region could also reconstitute C2′ activity, while various other proteins tried including BSA, immunoglobulins, and protein A could not. The apparent lack of a role for this basic region suggests that Myc's involvement in formation of this complex may be other than contacting DNA.

To further determine whether Myc was present in the analyzed complexes, the Myc preparations were incubated with a Myc-specific monoclonal antibody prior to EMSA. The probe used in this experiment

(SMS) contained a single site with the USE core sequence, CACGTG. The Myc-specific antibody eliminated both the C1 and C2 complexes and produced a prominent complex of slower mobility (Fig. 6). It is not clear from these data which of the two complexes was supershifted but the presence of one predominant shifted complex when antibody is present and two complexes in the absence of antibody argues that the Myc-specific antibody also completely disrupted one of the original complexes. There was no effect of a control monoclonal antibody on the formation of either the C1 or C2 complex. The Myc-specific antibody did not alter the C2′ complex, suggesting that Myc is not present in this complex.

From these experiments it can be concluded that the C1 complex is formed by Myc alone, that the C2 complex contains Myc and a 26-29 kd factor and that the C2′ complex contains a 40-50 kd factor but does not contain Myc. It is intriguing that the C2′ complex requires the presence of Myc for formation, but apparently does not contain Myc. Myc therefore appears capable of affecting the 40-50 kd factor's ability to form the C2′ complex without being a member of the complex. Whatever the mechanism, the increase in μE2 binding activity of the 40-50 kD factor appears to be Myc-specific since four different Myc proteins increased the amount of C2′ complex observed while several other proteins did not.

Max protein can be immunoprecipitated from avian and human cells and low stringency Southern analysis has suggested that a single Max gene or a small family of genes exist in other vertebrates as well (Blackwood and Eisenmann, *Science 251*:1211-1217 (1991)). It is possible that hamster and insect cells have an equivalent of Max. The recovery of a Max-like activity from insect cells is particularly interesting since no Myc homologs have been found in insects to date. Drosophila clearly uses B-HLH heterodimers to regulate aspects of development and the possibility remains that the natural partner for the 26-29 kD protein in insect cells is an as yet unidentified B-HLH protein which functions like Myc.

The presence of the 26-29 kD factor in these preparations might limit their usefulness for certain experiments. By pooling Myc containing fractions based on an EMSA assay, one may obtain fractions that contain only the C1 activity and that do not contain the C2 activity, although this modification reduces the final yield by approximately 80%.

Example 5

Selection of Binding Sites For Myc From Random DNA Sequences

In order to determine the optimal binding sites for the three complexes in the Myc preparations described above, a modification of a recently described technique for isolating preferred binding sites from large pools of randomized DNA sequences was used (Pollock and Treisman, *Nucl. Acids Res. 18*:6197-6204 (1990)). Briefly, a pool of double stranded oligonucleotides was created that consisted of 16 base flanking regions of defined sequence surrounding a 20 base region of completely random sequence. Each of the eukaryotic Myc preparations described above was mixed with this pool of sequences and the protein DNA complexes that formed were separated by EMSA. The DNA that ran at the position of the C1 or C2 (and comigrating C2') complexes was isolated, amplified by the polymerase chain reaction (PCR), and used in a second round of EMSA selection. Either seven (CHO preparation) or eight (baculovirus preparation) rounds of selection in total were performed before subcloning individual members of the selected sequences. As each round was expected to enrich for better binding sites, the final subcloned oligonucleotides were expected to contain high affinity binding sites for the C1, C2, and C2' complexes. In addition, such a procedure should give some indication of the relative stringency of selection for a given base at a particular position within the binding site consensus.

The selected sequences were placed in three separate groups for analysis (Fig. 7). Group I contains sequences that were selected by the C1

complex from either mammalian or baculovirus preparations. These sequences were pooled for analysis because with both preparations formation of the C1 complex requires only Myc protein, and because the two sets of sequences (that isolated with mammalian Myc and that isolated with baculovirus Myc) were similar to each other. Most of the selected sequences in this group contained the sequence CACGTG (21/27 of sequenced subclones). By aligning all of the sequences that contained this central core sequence, it was found that the sequences flanking this core were also nonrandom. A 12 base consensus sequence of GACCACGTGCTC [SEQ ID. No. 1] was determined for sites selected by the C1 complex (see table in Fig. 7 for frequencies at each position; for a base to be included in the consensus it had to be found in at least 10 out of the 21 sequences with a CACGTG core). The C2 complex from baculovirus preparations selected sequences similar to those selected by the C1 complex (Fig. 7, Group II). Most of these selected sequences also contained the CACGTG core (19/22). These sequences had similar flanking sequences adjacent to the core hexamer to those found with the C1 complex, although there was a slight preference for GCC over CTC in the 3′ flank (see table for Group II in Fig. 7).

As expected, complexes running at the position of C2 that were selected by the mammalian Myc preparations had a greater diversity of sequences (Fig. 7, Group III). Several sequences (8/36) contained the CACGTG core. These sequences were presumably selected by the mammalian C2 complex (comprised of Myc and the 26-29 kd factor) and demonstrated the same flank preferences as the C1 complex. Several other selected sites (9/36) contained a CAGCTG core sequence presumably selected by the C2′ complex. In addition, 8 of the 36 sequences were very AT rich, and many of the sequences in all three groups contained AT rich stretches. This enrichment for AT rich sequences might reflect a preference of Myc for these sequences, or instead might simply indicate a bias arising from the protocol used. It is interesting to note, however, that

in previous filter binding experiments, the mammalian Myc preparation has demonstrated a preference for binding AT rich sequences within various plasmids or lambda genomic DNA.

To confirm the validity of our site selection procedure a number of the selected sites individually by EMSA (Fig. 8) were tested. As expected, it was found that sequences containing the core CACGTG formed both the C1 and C2 complexes (Fig. 8, probe groups 1, 2, 5, and 6) while sequences containing the CAGCTG core formed only the C2′ complex (Fig. 8, probe groups 7 and 8). Note that the C2′ activity is only present in the CHO derived Myc preparations. No complex formed when selected sequences that did not contain a canonical E box core were tested (Fig. 8 probe groups 3, 4, 9, 10, and 11). These latter sequences, therefore, do not represent high affinity sites for proteins in the Myc preparations.

Example 6

Off Rate Of The C1 And C2 Complexes

Off-rates for the Myc containing complexes were measured as a means of comparing their affinities. The off-rate of the C2 complex formed on the USE probe was approximately 1-2 minutes (Fig. 9, baculovirus Myc; similar results were obtained with CHO Myc, data not shown). The C1 complex was not fully competed in this experiment using 250 fold excess of USE competitor. Although competition was not complete, the amount of C1 complex remaining at the earliest measurable timepoint ("0") was significantly less than the starting amount and virtually equal to the maximum competition achieved in these experiments. These data are indicative of an abundant weakly binding protein with an immeasurably fast off-rate. Therefore Myc alone appears to bind significantly more weakly than does Myc and the 26-29 kd factor.

Example 7

Identification of an Inhibitor of c-Myc C2 Complex Activity in Yeast Cells

Yeast host cells are transformed with plasmids carrying a c-Myc expression vector (host 'a'); or the c-Myc expression vector and a 26-29 kilodalton C2 complex protein identified as above (host 'b'). In addition all yeast strains are cotransformed with a plasmid that contains the coding sequence for β-galactosidase operably-linked to the CACGTG sequence motif as described above.

A lawn of each of the transformed yeast strains is spread on agar plates containing X-gal in the medium and small filter disks containing compound W, X, Y, or Z are placed on the lawns. The yeast are allowed to grow and the plates are monitored for colony growth and colony color by visual observation. Typical results from such an experiment are shown in Table 1.

Table 1: Identification of Inhibitors of C2 Complex Activity

Compound	Yeast	Colony Growth	Color from β-gal Assay with X-gal
none	a	+	White
	b	+	Blue
W	a	+	White
	b	+	White
X	a	-	-
	b	-	-
Y	a	+	White
	b	+	Blue
Z	a	+	Blue
	b	+	Blue

The results of the above table indicate that compound W prevents the induction of β-galactosidease in the 'b' host cells. Therefore, compound W is an inhibitor of C2 complex hetero-oligomer formation and an inhibitor of c-Myc biological activity. Compound X inhibits the growth of yeast *per se* and thus would not be a compound of interest.

Compound Y does not prevent induction of β-galactosidase activity in the 'b' host cells. Therefore, compound Y is not an inhibitor of C2 complex hetero-oligomer formation.

Compound Z shows an interesting effect of inducing β-galactosidase activity in the 'a' host cells which does not contain the C2 complex protein used in the 'b' hosts, rather than preventing hetero-oligomer formation. This suggests that compound Z may induce synthesis of a partner protein

which is not otherwise present in the yeast host cells or that it may be (or mimic) such a protein.

From these results, compound W would be identified as an inhibitor of C2 complex formation and/or DNA binding and thus of c-Myc transcriptional activity *in vivo*.

Example 8

Identification of an Inhibitor of c-Myc C2´ Complex Activity in Yeast Cells

Yeast host cells are transformed with two plasmids, each plasmid carrying a C2´ complex expression vector encoding at least one 40-50 kilodalton C2´ peptide (host 'a'); or the c-Myc expression vector in addition to the vectors encoding the C2´ complex proteins identified as above (host 'b'). In addition all yeast strains are cotransformed with a plasmid that contains the coding sequence for β-galactosidase operably-linked to the CAGCTG sequence motif as described above.

A lawn of each of the transformed yeast strains is spread on agar plates containing X-gal in the medium and small filter disks containing compound W, X, Y, or Z are placed on the lawns. The yeast are allowed to grow and the plates are monitored for colony growth and colony color by visual observation. Typical results from such an experiment are shown in Table 1.

Table 2: Identification of Inhibitors of C2´ Complex Activity

Compound	Yeast	Colony Growth	Color from β-gal Assay with X-gal
none	a	+	White
	b	+	Blue
W	a	+	White
	b	+	White
X	a	-	-
	b	-	-
Y	a	+	White
	b	+	Blue
Z	a	+	Blue
	b	+	Blue

The results of the above table indicate that compound W prevents the induction of β-galactosidase in the 'b' host cells. Therefore, compound W is an inhibitor of C2´ complex hetero-oligomer formation and an inhibitor of the c-Myc biological activity that is directed towards promoting such C2´ complex hetero-oligomer formation. Compound X inhibits the growth of yeast *per se* and thus would not be a compound of interest.

Compound Y does not prevent induction of β-galactosidase activity in the 'b' host cells. Therefore, compound Y is not an inhibitor of C2 complex hetero-oligomer formation.

Compound Z shows an interesting effect of inducing β-galactosidase activity in the 'a' host cells which does not contain the Myc protein used

in the 'b' hosts, rather than preventing hetero-oligomer formation. This suggests that compound Z may induce synthesis of a protein that can substitute for Myc in promoting formation of the C2´ complex which is not otherwise present in the yeast host cells or that it may be (or mimic) such a protein.

From these results, compound W would be identified as an inhibitor of C2´ complex formation and/or DNa binding activity and thus of c-Myc transcriptional activity *in vivo*.

All references cited herein are fully incorporated by reference. Having now fully described the invention, it will be understood by those with skill in the art that the scope may be performed within a wide and equivalent range of conditions, parameters and the like, without affecting the spirit or scope of the invention or any embodiment thereof.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Kingston, Robert E
Papoulas, Ophelia

(ii) TITLE OF INVENTION: C-MYC DNA BINDING PARTNERS, MOTIFS, SCREENING ASSAYS, AND USES THEREOF

(iii) NUMBER OF SEQUENCES: 101

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox
(B) STREET: 1225 Connecticut Avenue, N.W., Suite 300
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20036

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Cimbala, Michele A
(B) REGISTRATION NUMBER: 33,851
(C) REFERENCE/DOCKET NUMBER: 0609.3440004

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (202)833-7533
(B) TELEFAX: (202)833-8716

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GACCACGTGC TC 12

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GACCACGTGG TC 12

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AACAGTYCTG TT 12

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCGATATCTC GAGACGG 17

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGACGATCTA GATTCG 16

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATCTCTGCA GCAGCTGGCA GCAGCTGGCA GCAGCTGGCG 40

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCTGCAGT CATGTGGCGT CATGTGGCGT CATGTGGCAG 40

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATCTGCAGT CACGTGGCGT CACGTGGCGT CACGTGGCAG 40

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TCGACGTCGC AGCAGGTGCA G 21

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCGACCCCAC CAGCTGGCGA G 21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGCTTCGAAC ACCTGCAGCA GCTGGCAGGA AGCAGGCCTA 40

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGCTTTAAAA TCCCCACCAG CTGGCGAAGC AACAGGTGCA 40

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AATTGCGAAA CCCCTGGAAT ATTCCGACCT GGCAGCCTC 39

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCGACTTTAG ACCACGTGGT CCCCTCGA 28

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGACGATCTA GATTCGNNNN NNNNNNNNNN NNNNNNCCGT CTCGAGTATC GG 52

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

NNCANNTGNN 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCAGAATCTA CCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGGGCTACCA CGTGCTTATG 20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGACGAAAGC ACGTGCTCCG 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCACATGACC ACGTGCTCTG 20

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGCAGAGACA CGTGCCCTGG 20

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGCAAACCAC GTGTTATGTG 20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGACCACGTG CTCTTCGACT TG 22

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCACAATTTG TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGACAACATC GACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCCTGCATGA CCACGTGGAC C 21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCAAATATGA CCACGTGGTA C 21

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGACCACGTG CTCTTTTGTG 20

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATAAACT CCACGTGGTC C 21

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CGGGCACGTG CTCCTCGGAC TG 22

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GGTAGCAAAA AGCACGTGCC CG 22

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GGGGGATTTA AGCACGTGCT CC 22

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CACCTATTAA CCACGTGGTA C 21

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GACCACGCGG CATCCACGTG CCGT 24

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGGGACCACG TGCTCGGTTG 20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CACATATTAG ACCACGTGCT CC 22

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CGGCCACGTG CTCACTGTCT ACC 23

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGATGGACAG CTTCTTCCTG 20

(2) INFORMATION FOR SEQ Ib NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCAATCCCCC GCTGCTCGCC 20

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCCAAAAATG TACAGCTGTG CC 22

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CGGCCACGAG GTCATGAATG TGC 23

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCAGGCTGTA CGTGACTTGG 20

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CCGCAGTCCT GGTGCTCTGC 20

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CACTAAGAAA TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GGGGATTTAA GCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGGCCACGTG CCTTCTTTCT CCG 23

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CATAGTCGAG AGAGCACGTG CGAA 24

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CATAAGTCAG ACCACGTGGC CG 22

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CCCAACTAAG ACCACGTGGC CG 22

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CAGTCGAAGA GGCCACGTGG CGA 23

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGTAGGTTAT TCCCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CATAAATAGG CCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGAAAATGTA CCACGTGCTC C 21

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GGAACAGACC ACGTGGCTTG 20

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GTACCACGTG CTTTTTTGGC 20

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CAGTCCGAGG AGCACGTGCC CG 22

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CGGCCACGTG TCGAGCATGA GTC 23

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CGGCCACGTG CTCGTAAATT TGC 23

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GCGACAAAAT TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CGCAAAATCG ACCACGTGGT CC 22

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCATAAGTAA TACCACGTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GCAAAAAAAC CACGTGGTCC 20

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GGGGGCCGAA CTCCGTTGTC 20

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GGGGACCCGA TCTCTCGCTG 20

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CAATAATATT TCCTTTCCTG 20

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTCCACGCGG CATCCACGTG CCGT 24

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CGGCCACGTG CTCTATACAT GCC 23

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GGACCACGTG CTTATCTTTG 20

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CGACCACGTG TTCCGCTACT CG 22

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CGAGTAGCGA GCACGTGTTG C 21

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GCACCACGTG CTTACCATGT C 21

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GGACAAAAAG CACGTGCTAC 20

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GCAAAACTCC ACGTGGTCGG 20

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GGGCAAAAAC AACAGCTGTG CG 22

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GGGAAAGAGA TCAGCTGTGC G 21

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GGAGAATTGA ACAGCTGACC C 21

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GGGACAAACC AGTCAGCTGG CCG 23

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GGGCACAGCT GTTTAGTGGG 20

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

GGCAAGCGGA CAGCTGTTCC 20

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GGCATTGATC AGCTGTGTGG 20

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

GCAAAAACCA GCTGGTCCCC 20

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

CGCAAGTGTA ACAGCTGGTG C 21

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GGATGGTTTT TTTTTTGTAC 20

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GCATGATTTT CTTTTTGTCC 20

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

CAGAGTTTTT TTGAGCCCCC 20

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

GCAAAAAATA AAAATACATC 20

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GGCAAAAAAG TCAAAATACG 20

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GCACAATAAA AAACTTTGCG 20

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

CCATATGTTC ATTGTTGTCC 20

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

CACAAAAATT TAGTGTGTGC 20

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

CGGCCCCGTG CTCTAGCCCA TGC 23

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

CGGGGAAGTC CCAAGTGCCC C 21

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

CACAGGAACA TACACGGGCC CG 22

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GGGACGGGAT GATTGACGTG CCGT 24

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

CGCAAGCGAC GTCAGTCCTG 20

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CACCTACCAC TGATCGCGGC 20

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

GGACAAACAT CCCATTACCC 20

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

GGGGATGGAA CATCGCGCTG 20

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

CCAGTCGGGC CTAACCGGCC 20

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GGGAGCCATC GACGCCGGTG 20

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

CCATAGGGGA GTTGACAGCC 20

WHAT IS CLAIMED IS:

1. A method for the purification of Myc from a mammalian source, wherein said method comprises:

(a) growing mammalian cells capable of expressing c-Myc;

(b) inducing c-Myc expression in said cells;

(c) lysing the membrane of said mammalian cells and purifying nuclei therefrom;

(d) treating said nuclei in a buffer comprising DNase I;

(e) solubilizing said nuclei in a buffer comprising sodium dodecyl sulfate or urea at concentrations greater than 4 M and separating the nuclear pellet from the supernatant fraction;

(f) applying said supernatant fraction of step (e) to a DEAE Sepharose CL-6B column and eluting bound c-Myc from said DEAE Sepharose CL-6B column with a salt gradient;

(g) applying said c-Myc of step (f) to a FPLC Mono-Q column and eluting bound c-Myc with a salt gradient.

2. A method for the detection of C1 complexes in a sample, wherein said method comprises detecting DNA binding of c-Myc-containing homo-oligomers to the DNA motif 5'-CACGTG-3', in its double stranded DNA form.

3. A method for the detection of C2 complexes in a sample, wherein said method comprises detecting DNA binding of c-Myc-containing hetero-oligomers to the DNA motif 5'-CACGTG-3', in its double stranded DNA form.

4. A method for the detection of C2´ complexes in a sample, wherein said method comprises detecting c-Myc directed DNA binding to the DNA motif 5'-CAGCTG-3', in its double stranded DNA form.

5. A protein composition comprising at least one peptide capable of forming a C2 complex, wherein said peptide capable of forming a C2 complex is found in a 26-29 kD protein fraction purified from Chinese hamster ovary cells or baculovirus.

6. The protein composition of claim 5, wherein said protein composition is prepared from Chinese hamster ovary cells by a method comprising the steps of:

(a) growing said cells;

(b) lysing the membrane of said cells and purifying nuclei therefrom;

(c) treating said nuclei in a buffer comprising DNase I;

(d) solubilizing said nuclei in a buffer comprising sodium dodecyl sulfate or urea at concentrations greater than 4 M and separating the nuclear pellet from the supernatant fraction;

(e) applying said supernatant fraction of step (e) to a DEAE Sepharose CL-6B column and the bound C2 complex protein from said DEAE Sepharose CL-6B column with a salt gradient; and

(g) applying the eluted C2 complex protein of step (f) to a FPLC Mono-Q column and eluting bound C2 complex protein with a salt gradient.

7. The protein composition of claim 5, wherein said protein composition is prepared from baculovirus.

8. A protein composition comprising at least one peptide capable of forming a C2′ complex in the presence of c-Myc, wherein said peptide capable of forming a C2′ complex in the presence of c-Myc is found in a 40-50 kD protein fraction purified from CHO cells.

9. The protein composition of claim 8, wherein said protein composition is prepared from Chinese hamster ovary cells by a method comprising the steps of:

(a) growing said cells;

(b) lysing the membrane of said cells and purifying nuclei therefrom;

(c) treating said nuclei in a buffer comprising DNase I;

(d) solubilizing said nuclei in a buffer comprising sodium dodecyl sulfate or urea at concentrations greater than 4 M and separating the nuclear pellet from the supernatant fraction;

(e) applying said supernatant fraction of step (e) to a DEAE Sepharose CL-6B column and the bound C2′ complex protein from said DEAE Sepharose CL-6B column with a salt gradient; and

(g) applying the eluted C2′ complex protein of step (f) to a FPLC Mono-Q column and eluting bound C2′ complex protein with a salt gradient.

10. A method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity, wherein said method comprises detecting the ability of said compound to inhibit C1 complex formation, C2 complex formation or C2′ complex formation.

11. The method of claim 10, wherein said complex formation is C1 complex formation.

12. The method of claim 10, wherein said complex formation is C2 complex formation.

13. The method of claim 10, wherein said complex formation is C2′ complex formation.

14. A method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity, wherein said method comprises detecting the ability of said compound to inhibit C1 complex DNA binding, C2 complex DNA binding, or C2′ complex DNA binding.

15. The method of claim 14, wherein said DNA binding is C1 complex DNA binding.

16. The method of claim 15, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CACGTG-3'.

17. The method of claim 15, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CATGTG-3'.

18. The method of claim 14, wherein said DNA binding is C2 complex DNA binding.

19. The method of claim 18, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CACGTG-3'.

20. The method of claim 18, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CATGTG-3'.

21. The method of claim 14, wherein said DNA binding is C2´ complex DNA binding.

22. The method of claim 21, wherein said DNA binding is to an oligonucleotide comprising the sequence 5'-CAGCTG-3'.

23. A method for the purification of a peptide capable of forming a C2 or C2´ complex, or a mixture of such peptides from a crude preparation, wherein said method comprises extraction of Chinese hamster ovary cells and assay of said peptide by detection of the ability of said peptide to form said C2 or said C2´ complex.

24. A method for identifying and classifying a compound as an inhibitor of c-myc hetero-oligomer DNA binding wherein said method comprises evaluating the ability of said compound to alter expression of a reporter gene in a host cell, wherein expression of said reporter gene is operably-linked to DNA binding by said hetero-oligomer to an oligonucleotide comprising the sequence 5'-CACGTG-3'.

25. The method of claim 24, wherein expression of said reporter gene induces a phenotypic change in a host cell.

26. The method of claim 24, wherein said reporter gene is *lacZ*.

27. The method of claim 24, wherein said reporter gene is.CAT.

28. The method of claim 24, wherein said reporter gene is *LEU2*.

29. The method of claim 24, wherein said phenotypic change is detected by visual inspection of the host cell.

30. The method of claim 24, wherein said host is *S. cerevisiae*.

31. The method of claim 24, wherein said host is a mammalian cell.

32. A method for identifying and classifying a compound as an inhibitor of c-Myc-directed C2´ hetero-oligomer DNA binding wherein said method comprises evaluating the ability of said compound to alter expression of a reporter gene in a host cell, wherein expression of said reporter gene is operably-linked to DNA binding by said hetero-oligomer to an oligonucleotide comprising the sequence 5'-CAGCTG-3'.

33. The method of claim 32, wherein expression of said reporter gene induces a phenotypic change in a host cell.

34. The method of claim 32, wherein said reporter gene is *lacZ*.

35. The method of claim 32, wherein said reporter gene is CAT.

36. The method of claim 32, wherein said reporter gene is *LEU2*.

37. The method of claim 32, wherein said phenotypic change is detected by visual inspection of the host cell.

38. The method of claim 32, wherein said host is *S. cerevisiae*.

39. The method of claim 32, wherein said host is a mammalian cell.

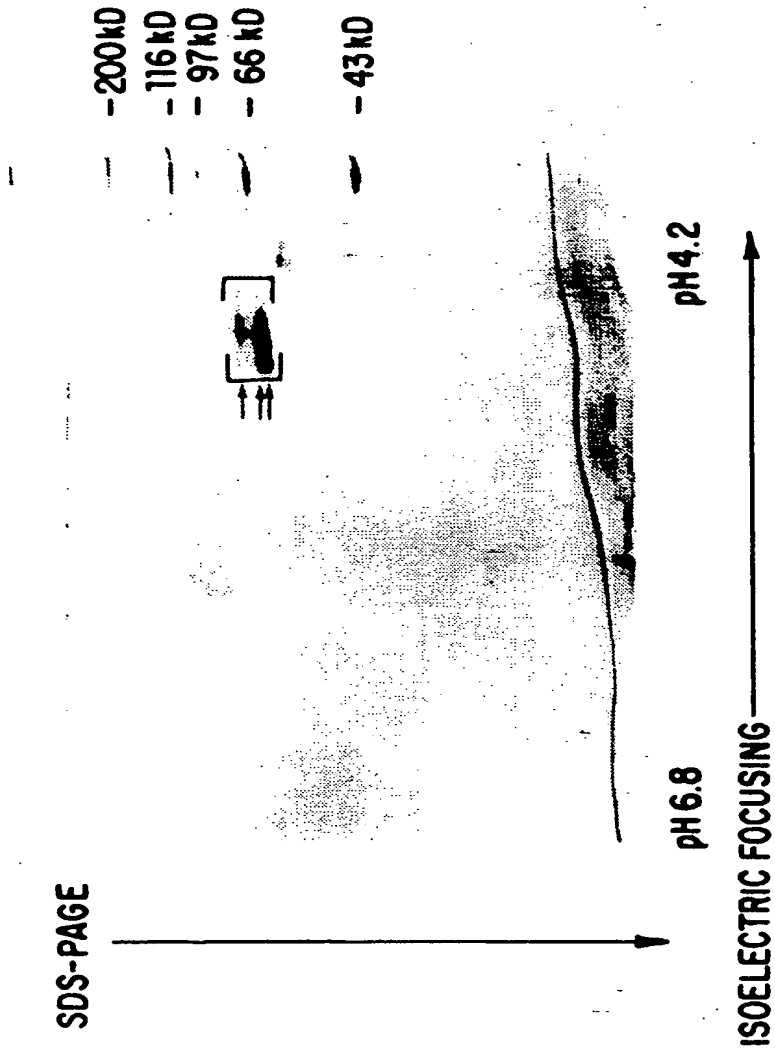


FIG. 1A

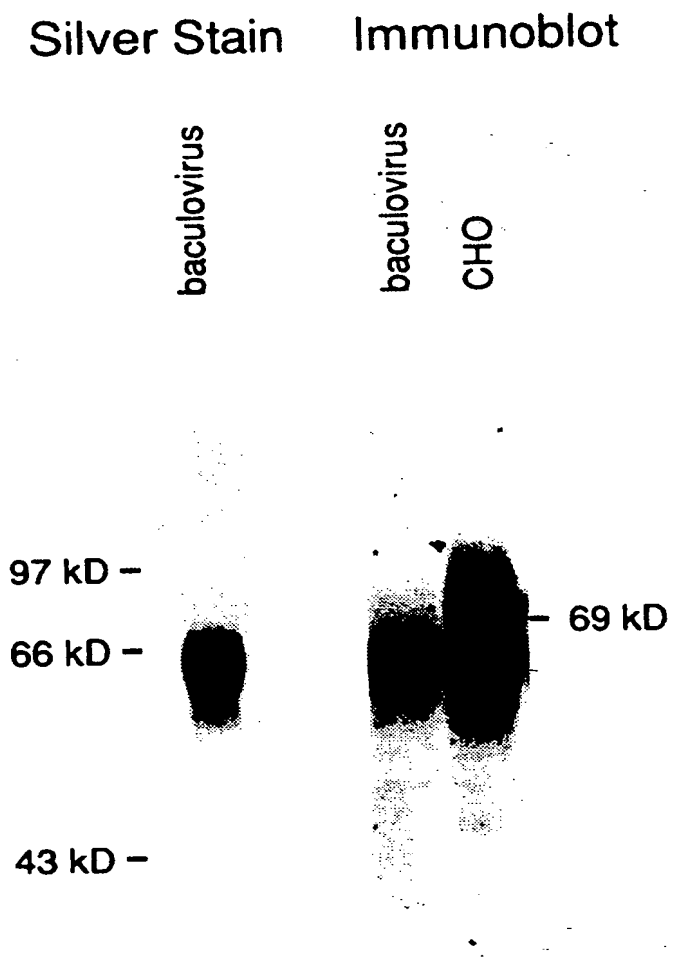
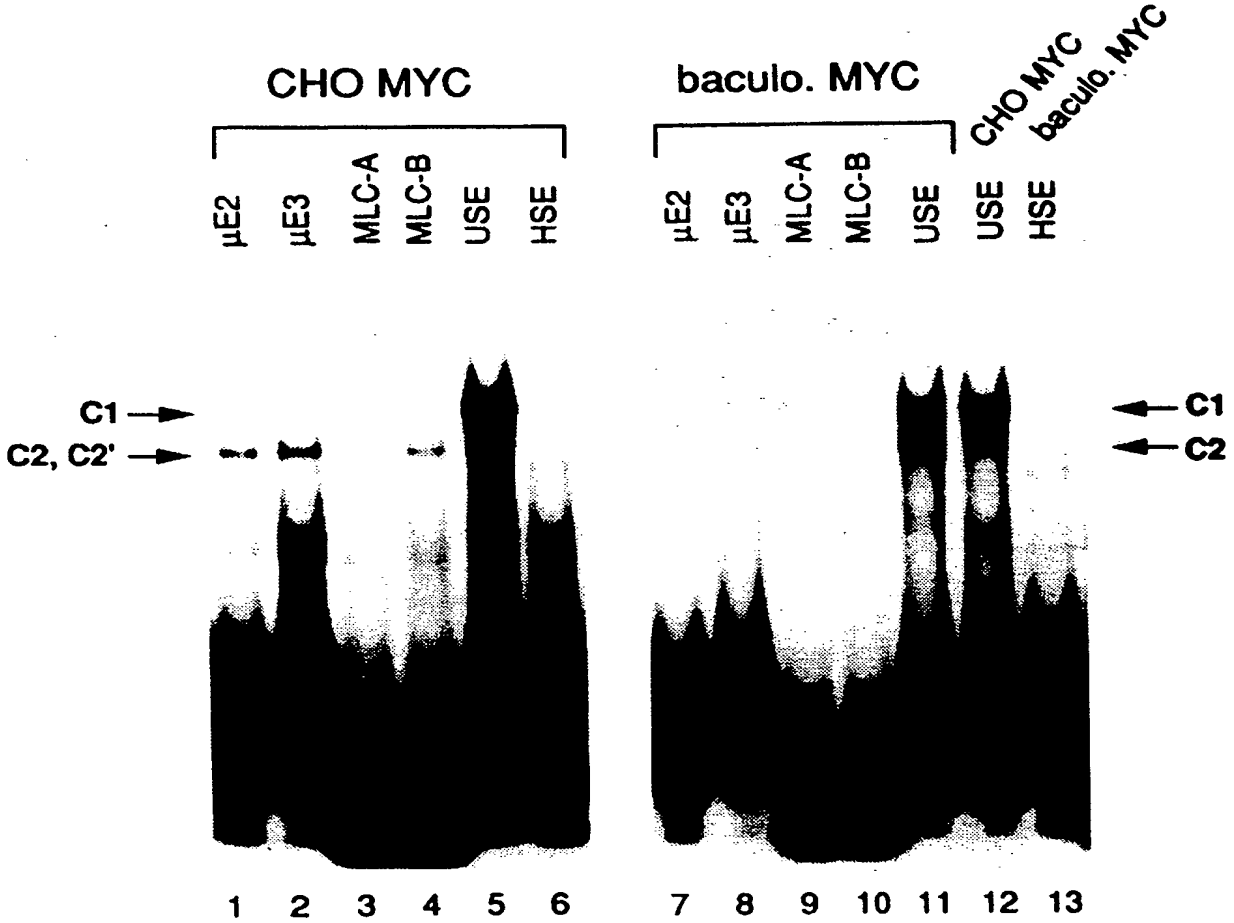


FIG. 1B



E-Box Probes

MLC-B (µE2)	CAGCTG
µE3	CATGTG
USE	CACGTG
MLC-A (kE2)	CAGGTG

FIG. 2

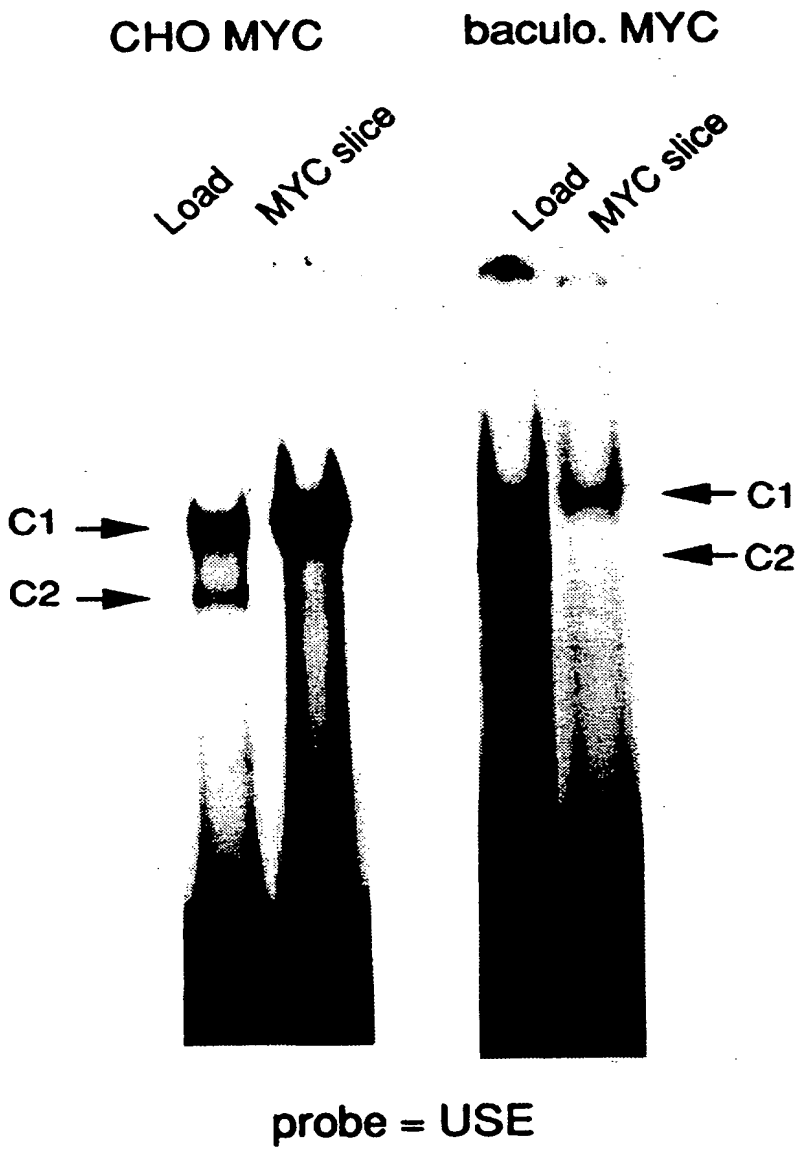


FIG. 3

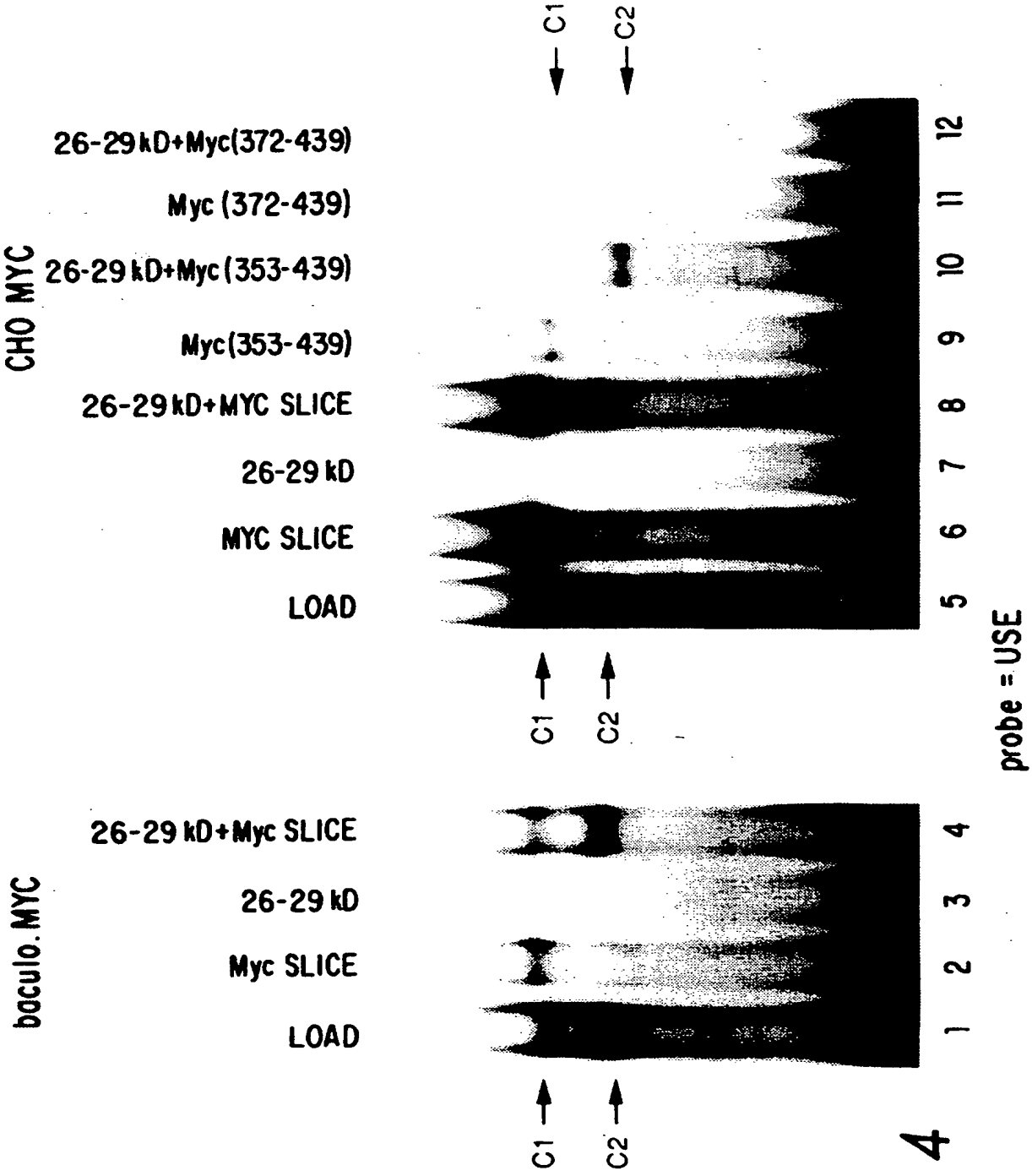


FIG. 4

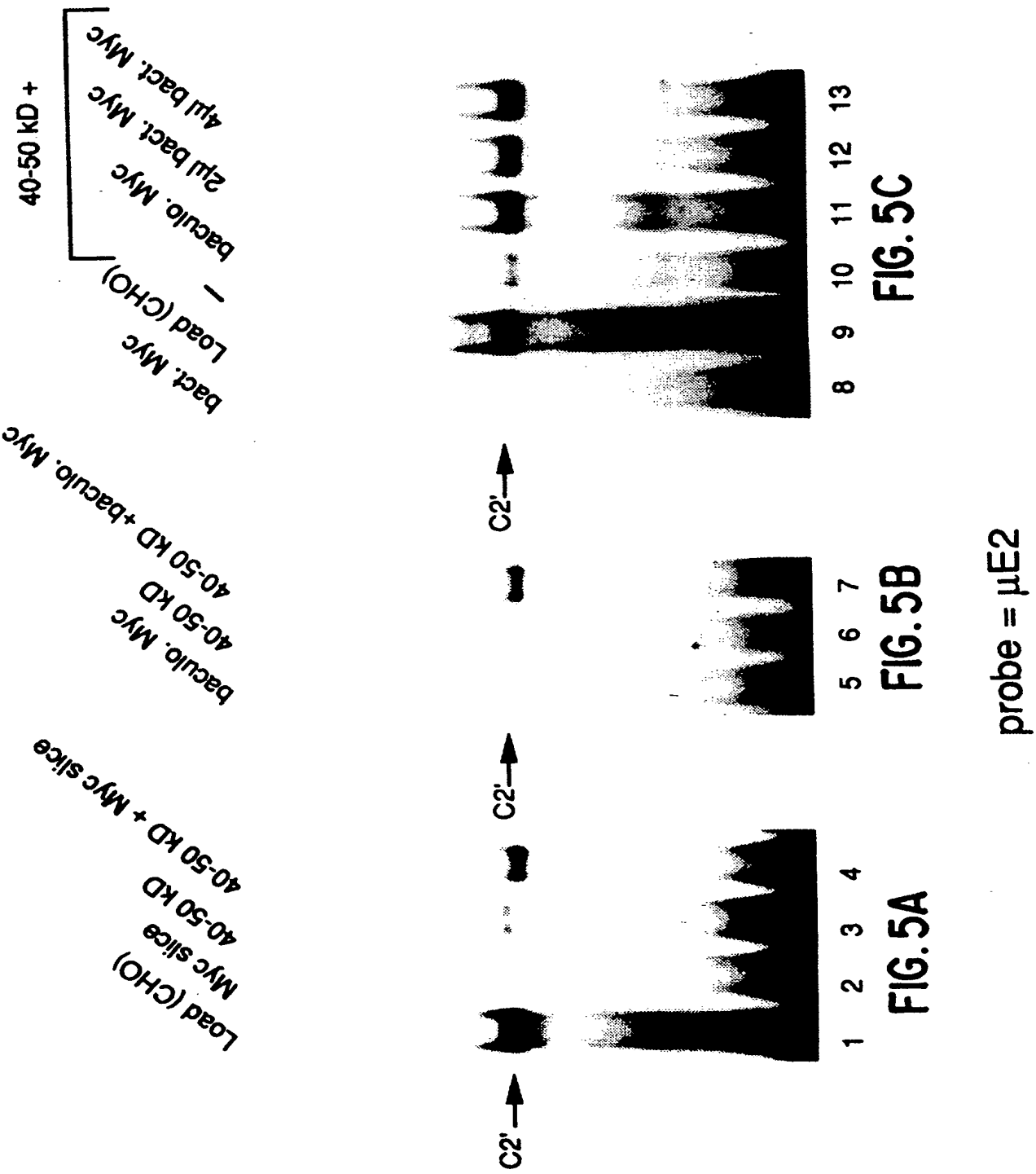


FIG. 5A

FIG. 5B

FIG. 5C

SUBSTITUTE SHEET

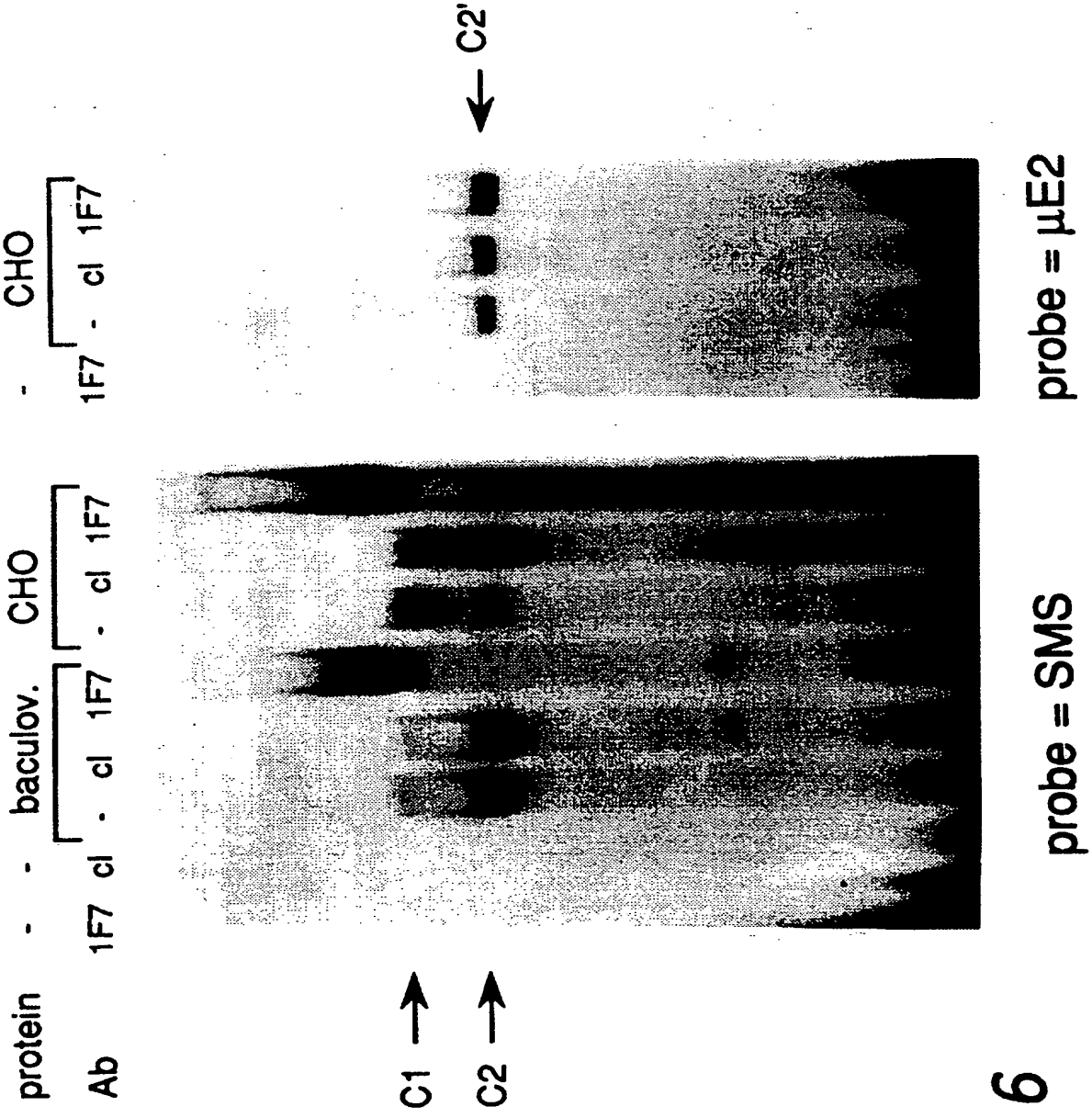


FIG. 6

GROUP I

GCAGAATCTAC CACGTG CTCC
GGGGCTAC CACGTG CTTATG
GGACGAAAG CACGTG CTCCG
GCACATGAC CACGTG CTCTG 1
GGCAGAGA CACGTG CCCTGG
GGCAAAC CACGTG TTATGTG *
CGAC CACGTG CTCTTCGACTTG
GCACAATTTGTAC CACGTG GCCG
GGACAACATCGAC CACGTG GCCG
GCCTGCATGAC CACGTG GACC
GCAAATATGAC CACGTG GTAC
GGAC CACGTG CTCTTTTGTG
GGCATAAACTC CACGTG GTCC
CGGG CACGTG CTCCTCGGACTG
GGTAGCAAAAAG CACGTG CCCG 2
GGGGGATTTAAG CACGTG CTCC
CACCTATTAAC CACGTG GTAC
GACCACGCGGCATC CACGTG CCGT
GGGGAC CACGTG CTCGGTTG
CACATATTAGAC CACGTG CTCC
CGGC CACGTG CTCACTGTCTACC

GGATGGA CAGCTT CTTCCTG
GCAATCCC CCGCTG CTCGCC 3
GCCAAAAATGTA CAGCTG TGCC
CGGC CACGAG GTCATGAATGTGC
GCAGGCTG TACGTG ACTTGG
CCGCAGTC CTGGTG CTCTGC

GROUP II

CACTAAGAAATAC CACGTG GCCG
GGGGATTTAAG CACGTG CTCC
CGGC CACGTG CCTTCTTTCTCCG
CATAGTCGAGAGAG CACGTG CGAA
CATAAGTCAGAC CACGTG GCCG
CCCAACTAAGAC CACGTG GCCG *
CAGTCGAAGAGGC CACGTG GCGA
CGTAGGTTATTCC CACGTG GCCG
CATAAATAGGC CACGTG CTCC *
GGAAAATGTAC CACGTG CTCC
GGAACAGAC CACGTG GCTTG
GTAC CACGTG CTTTTTTGGC
CAGTCCGAGGAG CACGTG CCCG
CGGC CACGTG TCGAGCATGAGTC
CGGC CACGTG CTCGTAAATTTGC
GCGACAAAATTAC CACGTG GCCG
CGCAAAATCGAC CACGTG GTCC
GCATAAGTAATAC CACGTG GCCG
GCAAAAAAAC CACGTG GTCC

GGGGGCGGAACTCCGTTGTC 4
GGGGACCCGATCTCTCGCTG
CAATAATATTTCCTTTCCTG

	-3	-2	-1	CACGTG	1	2	3
C	1	0	16		14	5	16
G	10	3	4		6	0	1
A	7	16	1		0	1	3
T	3	2	0		1	15	1

	-3	-2	-1	CACGTG	1	2	3
C	0	1	16		8	11	13
G	11	5	3		10	1	2
A	2	13	0		0	0	1
T	6	0	0		1	7	3

FIG.7A

GROUP III μE2 CORE (CAGCTG)

USE CORE (CACGTG)

```
GTCCACGCGGCATC CACGTG CCGT
        CGGC CACGTG CTCTATACATGCC
        GGAC CACGTG CTTATCTTTG   5
        CGAC CACGTG TTCCGCTACTCG  *
   CGAGTAGCGAG CACGTG TTGC
        GCAC CACGTG CTTACCATGTC
    GGACAAAAAG CACGTG CTAC
     GCAAAACTC CACGTG GTCGG      6
```

	-3	-2	-1	CACGTG	1	2	3
C	2	0	6		5	1	3
G	4	1	2		1	0	2
A	2	5	0		0	0	1
T	0	2	0		2	7	2

```
  GGGCAAAAACAA CAGCTG TGCG
   GGGAAAGAGAT CAGCTG TGCG
   GGAGAATTGAA CAGCTG ACCC
  GGGACAAACCAGT CAGCTG GCCG
        GGGCA CAGCTG TTTAGTGGG
   GGCAAGCGGA CAGCTG TTCC     7
    GGCATTGAT CAGCTG TGTGG    *
     GCAAAAAC CAGCTG GTCCCC   8
  CGCAAGTGTAA CAGCTG GTGC
```

AT-RICH

```
GGATGGTTTTTTTTTTGTAC
GCATGATTTTCTTTTTGTCC
CAGAGTTTTTTTGAGCCCCC
GCAAAAAATAAAAATACATC
GGCAAAAAAGTCAAAATACG
GCACAATAAAAAACTTTGCG      9
CCATATGTTCATTGTTGTCC
CACAAAAATTTAGTGTGTGC      10
```

OTHER

```
          CGGC CCCGTG CTCTAGCCCATGC
    CGGGGAAGTCC CAAGTG CCCC     *
   CACAGGAACATA CACGGG CCCG
  GGGACGGGATGATT GACGTG CCGT
        CGCAAGC GACGTC AGTCCTG
  CACCTACCACTGATCGCGGC          11
  GGACAAACATCCCATTACCC          *
  GGGGATGGAACATCGCGCTG          *
  CCAGTCGGGCCTAACCGGCC
  GGGAGCCATCGACGCCGGTG
  CCATAGGGGAGTTGACAGCC
```

FIG. 7B

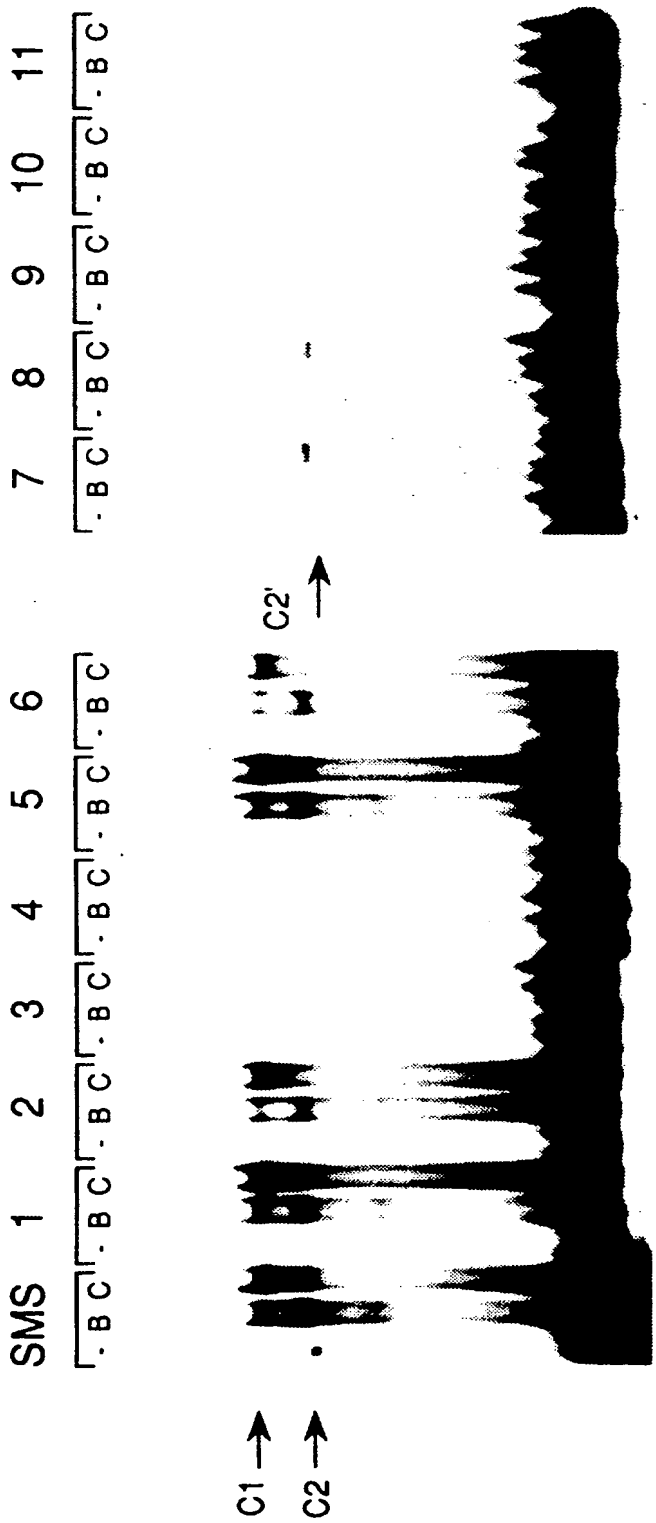


FIG. 8

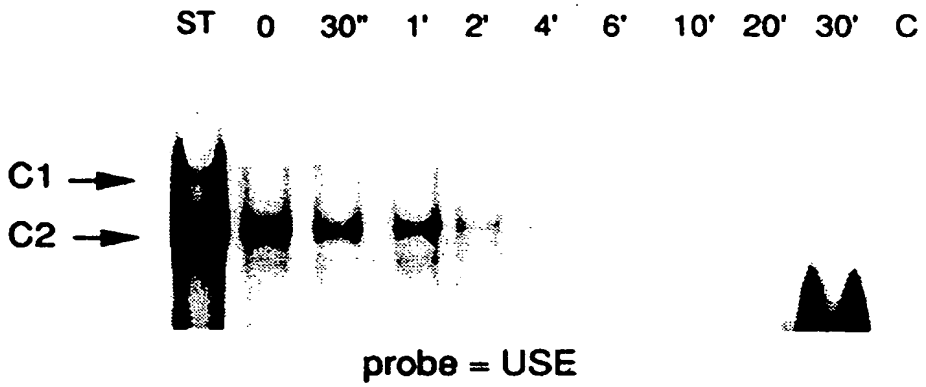


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :A23J 1/00; C07K 3/00; C12Q 1/68; G01N 1/00
US CL :530/412, 417; 435/6; 530/350; 424/2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/412, 417; 435/6; 530/350; 424/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Prior art disclosure of priority document.

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS
DIALOG (Biochem)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCIENCE, Volume 250, issued 23 November 1990, T.K. Blackwell, *et alii*, "Sequence-Specific DNA Binding by the c-Myc Protein", pages 1149-1151; see pages 1151, column 1.	2-4, 14-24, 32-39
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES UNITED STATES OF AMERICA, Volume 81, issued December 1984, G. Ramsey *et alii*, "The Protein Encoded by the Human Proto-Oncogene c-*myc*", pages 7742-7746, see abstract; page 7744, second column; and page 7745 first column.	10, 11
X	SCIENCE, Volume 225, issued 17 August 1984, H. Persson *et alii*, "Nuclear Localization and DNA Binding Properties of a Protein Expressed by Human c-*myc* Oncogene", pages 718-721, see page 719, first column.	8,9
X	MOLECULAR AND CELLULAR BIOLOGY, Volume 5, number 3, issued March 1985, R.A. Watt *et alii*, "Expression and Characterization of the Human c-*myc* DNA Binding Protein", pages 448-456, see page 450, first column; page 451, second column; and page 455, first column.	1-4, 23

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 09 DECEMBER 1992	Date of mailing of the international search report 11 JAN 1993
Name and mailing address of the ISA/ Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer MARGARET MOSKOWITZ Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, UNITED STATES OF AMERICA, Volume 83, issued August 1986, F. M. Wurm, "Inducible overproduction of the mouse c-*myc* protein in mammalian cells", pages 5414-5418, see page 5416, second column.	24, 25, 29, 31, 32, 37, 39
X	NATURE, Volume 296, issued 18 March 1982, P. Donner, *et alii*, "Nuclear Localization and DNA Binding of the Transforming Gene Product of Avian Myelocytomatosis Virus", pages 262-266, see page 262, second column; and page 263.	5-9
X	ANN. REV. GENET., Volume 120, issued 1986, M.D. Cole, "The *myc* Oncogene: Its Role in Transformation and Differentiation", pages 361-384; see the entire document.	10-21, 24-39

Form PCT/ISA/210 (continuation of second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1 and 23, drawn to a method of purifying Myc, classified class 530, subclass 412 and 417, and claims 5-9, drawn to a protein classified in class 530, subclass 350.

Group II, claims 2-4, drawn to a method of detection, classified in class 435, subclass 6.

Group III, claims 10-22 and 24-39, drawn to a method of identifying and/or classifying compounds, said claims are classified in class 424, subclass 2.

The inventions listed as Groups I and II do not meet the requirements for Unity of Invention for the following reasons: The invention of Group I is considered to form an inventive concept wherein said first method is directed to a purification of Myc and the product is at least one peptide capable of forming a C2 complex and that said peptide is found in a 26-29 kD protein fraction purified from Chinese hamster ovary cells or baculovirus. The invention of Group II is drawn to different methodologies, requiring different method steps and resulting in a different end product. In particular, the invention of Group II is directed to the detection of C1 complexes.

Additionally, the peptide of claim 5, "a protein composition comprising at least one peptide capable of forming a C2 complex "is admittedly and agreeably old in the art. As can be seen on page 1, bridging to page 2 of the disclosure, myc has long been known in the art as well as its sequence. Also, the prior art of record clearly demonstrates that myc has been expressed under recombinant conditions. Resultantly, the peptide of claims 5-9 cannot be considered as a special technical feature. Lacking such status, the holding of a lack of unity is justified.

The inventions listed as Groups I and III do not meet the requirements for Unity of Invention for the following reasons: As shown *supra*, the invention of Group I is comprised of a first method of making and a product made where said first method is directed to the purification of Myc and the product is at least one peptide capable of forming a C2 complex and that said peptide is found in a 26-29 kD protein fraction purified from Chinese hamster ovary cells or baculovirus. The invention of Group III is directed to "[a] method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity" (claims 10-22) and "[a] method for identifying and classifying a compound as an inhibitor of c-myc hereto-oligomer DNA binding" (claims 24-39). As can be clearly seen, the inventions of each group are drawn to different methodologies, requiring different method steps and resulting in different end products.

The inventions listed as Groups II and III do not meet the requirements for Unity of Invention for the following reasons: The inventions of Group II is drawn to a method of detection of C1 complexes while the invention of Group III is drawn to "[a] method for objectively classifying compounds, including human pharmaceuticals, as inhibitors of c-Myc activity" (claims 10-22) and "[a] method for identifying and classifying a compound as an inhibitor of c-myc hetero-oligomer DNA binding" (claims 24-39). Clearly said Groups are drawn to methodologies that each require different method steps and result in different end products.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/08603

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims. (Telephone Practice)

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any addtional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention frist mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.